

Gabriele Filosofi

Appunti integrati su

MANUALE DI SCIENZA COGNITIVA

E.Pessa, M.Pietronilla Penna

RETI NEURALI E PROCESSI COGNITIVI

E.Pessa

MODELLI QUANTISTICI DELLA MEMORIA

E.Pessa, A.Ferretti

MIND, MATTER, AND QUANTUM MECHANICS

Henry P. Stapp

NEURAL NETWORKS FOR PATTERN RECOGNITION

C.M.Bishop

IL PROBLEMA DELLA CONOSCENZA

- A partire dal '600 la conoscenza, facendosi "scienza", è in grado di produrre effetti sistematici sul mondo; a partire dal 1970 si cerca di creare una nuova scienza, la **Scienza Cognitiva (CS)**.
- Nello studio della conoscenza distinguiamo l'impostazione platonica dall'impostazione aristotelica.

Secondo l'**impostazione platonica** la conoscenza è rappresentazione della realtà immutabile che esiste oltre i fenomeni reali. Tale conoscenza, come anche la realtà "vera" che essa descrive, è strutturata secondo relazioni logico-matematiche. Resta il problema di definire le relazione tra conoscenza e mondo sensibile.

Secondo l'**impostazione aristotelica** quella fenomenica è l'unica realtà. La conoscenza estrae dai fenomeni i principi generali del loro divenire, che però non esistono indipendentemente dai fenomeni stessi. Resta il problema di chiarire le modalità di questa "estrazione".

LOGICA MATEMATICA

- L'impostazione platonica ha ispirato la teoria filosofica della conoscenza e la teoria Matematica.
- Dalla *Logica Filosofica* nacque in quel periodo la **Logica Matematica** (o **Calcolo Proposizionale**) ad opera di **George Boole** e **Augustus De Morgan**.
- Per i **formalisti**, con a capo **David Hilbert**, ogni teoria matematica è un **sistema formale**, costruzione di sequenze simboliche (o **proposizioni**, **P**) primitive (**assiomi**), o derivate (**teoremi**) per mezzo di regole di produzione (o **regole di inferenza**).
- Def. Il sistema è **consistente** \Leftrightarrow non è possibile che siano veri entrambe i teoremi **P** e $\sim P$.
- Def. Il sistema è **completo** \Leftrightarrow è possibile in linea di principio decidere la verità o falsità di qualsiasi proposizione esprimibile all'interno del sistema.
- Negli anni '80 **Georg Cantor** aveva definito una teoria dei vari tipi di infinito, nota come **Teoria degli Insiemi**.
- **Gottlob Frege** a Jena e **Giuseppe Peano** a Torino abbinarono la logica formale alla teoria degli insiemi e dei numeri. Nel 1884 Frege aveva pensato di esprimere i numeri come insiemi, così l'**insieme 3** di Frege è l'insieme di tutti gli insiemi di tre elementi, ecc.
- Hilbert a Gottinga rielaborò la geometria Euclidea in termini più rigorosi. In quel periodo vennero alla luce una serie di paradossi, tra i quali, nel 1902, il **paradosso di Russell** su "l'insieme di tutti gli insiemi che non appartengono a sé stessi". Gli sforzi della logica formale culminarono nel lavoro monumentale di **Bertrand Russel** e **Alfred North Whitehead**, *Principia Mathematica* (1913), dove si costruiva un sistema altamente formalizzato di assiomi e regole procedurali tale da

non consentire ragionamenti paradossali. Hilbert lanciò una sfida alla comunità dei matematici: dimostrare che il sistema definito nei *Principia Mathematica* era sia *coerente*, sia *completo*. In realtà il progetto di Hilbert era ancora più ambizioso: trovare, per ogni area ben definita della matematica, un elenco di assiomi e regole procedurali capace di includere tutte le forme di ragionamento matematico corretto appropriate a tale area. Ma nel 1931 **Kurt Gödel** demolì il programma di Hilbert, mettendo in luce i limiti della Logica Matematica.

- Il **Teorema di Incompletezza di Gödel** compare come la Proposizione VI del suo scritto “*Sulle proposizioni formalmente indecidibili dei Principia Mathematica e di sistemi affini*” (1931):

“Ad ogni classe k di formule che sia ω -coerente e ricorsiva corrispondono *segnidi-classe* ricorsivi r tali che né $v \text{ Gen } r$ né $\text{Neg } (v \text{ Gen } r)$ appartengono a $\text{Flg } (k)$ (dove v è la *variabile libera* di r)”

che può essere tradotta

“Ogni sistema assiomatico consistente e adeguato a esprimere un insieme di asserzioni sufficientemente ampio sui numeri interi è incompleto, cioè possiede teoremi veri che non possono essere provati”

oppure

“Tutte le assiomatizzazioni coerenti dell’aritmetica contengono proposizioni indecidibili”

- La dimostrazione di Gödel consisteva di due parti:
 - 1) mostrare come poter codificare le regole di produzione e l’uso degli assiomi in operazioni aritmetiche, sicché ogni *funzione proposizionale* sintatticamente corretta tipo $P(m) = \sim \exists p, q, r \mid (p+1)^{m+3} + (q+1)^{m+3} = (r+1)^{m+3}$ (teorema di Fermat) e ogni *dimostrazione* Π , come sequenza di proposizioni, sono associabili a un numero naturale ($n \leftrightarrow P_n \quad p \leftrightarrow \Pi_p$)
 - 2) considerare la funzione proposizionale $P_k(m) = \sim \exists p \mid \Pi_p \text{ dimostra } P_m(m)$. Ora, posto $m=k$, abbiamo

$$P_k(k) = \sim \exists p \mid \Pi_p \text{ dimostra } P_k(k)$$

Se $P_k(k)$ fosse falsa avremmo che esiste una dimostrazione che porta ad essa, ma ciò è impossibile perché una dimostrazione è una sequenza di proposizioni vere per definizione, che termina con la proposizione da dimostrare. Quindi $P_k(k)$ dev’essere vera, ma in tal caso *non esiste per essa una dimostrazione all’interno del sistema formale*.

RIDUZIONISMO E FISICALISMO

- L’impostazione aristotelica ha portato agli sviluppi della Fisica post-galileiana.
- Il primato della Fisica tra le scienze ha portato molti a credere che essa fosse l’unica scienza, alla quale tutte le altre sarebbero prima o poi ricondotte

(*fisicalismo*). Tale concezione stimola all'unificazione delle scienze, ma storicamente ha finito per privilegiare la *Meccanica Classica (CM)*, i cui limiti, oggi ben noti, hanno portato a una giusta reazione che rivendica l'autonomia delle altre discipline scientifiche (Psicologia, Biologia, Economia, ecc.).

- Il fisicalismo è un caso particolare di *riduzionismo*. Secondo il *riduzionismo ontologico* oggetti e processi di un certo livello sono *costituiti* da oggetti e processi di un livello inferiore; secondo il *riduzionismo esplicativo* le proprietà di oggetti e processi di un certo livello sono *spiegabili* in termini delle proprietà di oggetti e processi di un livello inferiore
- Il riduzionismo ha perso molto della sua forza, anche perché in seno alla stessa Fisica vi è una miriade di fenomeni in cui a livello macroscopico emergono proprietà nuove, che a loro volta condizionano il comportamento dei costituenti microscopici (p.es. un cristallo, superconduttività, superfluidità)
- La CS, diversamente dalla Fisica, non ha un numero limitato di grandezze osservabili e intuitive a partire dalle quali ricavare le altre, come lunghezza, tempo, temperatura, ecc. specialmente in contesti come il problem solving, ecc.
- Secondo il concetto di *emergenza* proposto nel 1923 da *Conwy Lloyd Morgan*, quando la complessità di un sistema collettivo supera una certa soglia, si ha automaticamente la comparsa di proprietà globali, imprevedibili a partire dalla sola conoscenza delle proprietà dei costituenti
- Negli anni '40 il biologo *Ludwig Von Bertalanffy* definì la *Teoria Generale dei Sistemi*, studio dei sistemi complessi omeostatici che scambiano energia, materia e informazione con l'ambiente. Egli si avvale della *Teoria Matematica dei Sistemi Dinamici*, fondata da *Poincaré* e *Ljapunov* alla fine dell'800 nel contesto dell'Astronomia e della Meccanica.
- Oggi è l'*impostazione sistemica* che porta avanti lo studio sulla formazione dei sistemi complessi, l'interazione organismo-ambiente, la distinzione tra sistemi chiusi e aperti, l'incertezza e il caos, il ruolo dell'osservatore. Tra gli sviluppi teorici basati su questa impostazione la *Teoria delle Strutture Dissipative* di Prigogine, la *Sinergetica* di Haken, la *Soft System Methodology* di Checkland.

I PRINCIPI DELLA FISICA CLASSICA

- Il meccanicismo ha dominato il pensiero scientifico occidentale dai tempi di Galilei e Newton a oggi. La *CM* fondata da Newton si occupa di sistemi semplici, masse puntiformi, pendoli, corpi rigidi, ed è basata sui tre *principi della dinamica*:
 - 1) *ogni corpo non sottoposto ad azioni esterne persiste nel suo stato di quiete o di moto rettilineo uniforme*
 - 2) *in un sistema di riferimento inerziale l'accelerazione di un corpo puntiforme prodotta da forze applicate è direttamente proporzionale alla somma delle forze stesse ed è inversamente proporzionale alla massa del corpo*
 - 3) *due punti materiali isolati interagiscono mutualmente con forze uguali e opposte*

Fu presto evidente la limitatezza degli strumenti matematici nel ricavare, a partire dai tre principi, previsioni esplicite dell'evoluzione dinamica di sistemi a poche parti (problema dei tre corpi) e a molte parti interagenti (liquidi e gas). L'attenzione dei fisici si è spostata quindi sugli **stati di equilibrio**, ricavabili senza dover risolvere il sistema completo, con i metodi della **Teoria Qualitativa delle Equazioni Differenziali**, iniziata da Poincaré e Ljapunov. Per i sistemi a molte particelle si è introdotta la distinzione tra livelli di descrizione *microscopico* e *macroscopico*, quest'ultimo nei termini di poche grandezze globali, ottenute come medie statistiche sulle variabili microscopiche (p.es. P , T di un gas).

- La **Termodinamica Classica (CT)** associa a ciascuno stato di equilibrio una funzione di stato, l'**entropia termodinamica**. I tre principi della CT sono
 - 1) *Per un sistema termodinamico S esiste una funzione di stato, energia interna U , la cui variazione quando il sistema passa da uno stato 1 a uno stato 2 dipende solo dagli stati iniziale e finale; tale variazione è pari alla energia scambiata con l'esterno tramite il flusso di calore e il lavoro.*
 - 2) *E' impossibile realizzare una trasformazione il cui risultato finale sia solamente quello di trasformare in lavoro il calore estratto da una sorgente termica (Lord Kelvin); oppure, è impossibile realizzare una trasformazione il cui risultato finale sia solamente quello di trasferire calore da un corpo ad un altro a temperatura maggiore del primo (Clausius)*
 - 3) *Al tendere della temperatura allo zero assoluto, per qualsiasi trasformazione isoterma di un sistema che comporti solo fasi il cui stato dipende solo dalla temperatura, pressione e composizione, la variazione di entropia tende a zero; oppure, non è possibile, neanche per mezzo di un esperimento ideale, portare un sistema allo zero assoluto in un numero finito di operazioni (Nernst).*
- Helmholtz nel 1847 formulò il **principio della conservazione dell'energia**, di cui il 1° principio della termodinamica è un caso particolare, e che si esprime dicendo che *l'energia richiesta per far passare un sistema da uno stato a un altro non dipende dagli stati intermedi attraversati dal sistema, né dalla forma in cui l'energia è scambiata con l'esterno.*
- Le ricerche di Gibbs e Boltzmann portarono alla comprensione del fatto che la corrispondenza tra **macrostato** e **microstato** non è univoca, nel senso che a uno stesso macrostato possono corrispondere molti microstati, il numero W dei quali è una misura del disordine termodinamico, o **entropia statistica S** , associato al macrostato ($S = k_B \log W$). L'evoluzione del sistema chiuso tende spontaneamente allo stato di equilibrio di massima entropia, infatti si dimostra che se tutti i microstati hanno uguale probabilità a priori, lo stato di massima entropia è quello più probabile.
- L'esistenza di processi di crescita, differenziazione, cognitivi, nei quali si osserva un indubbia riduzione del disordine non deve lasciarci perplessi.

ASSOCIAZIONISMO

- Le teorie associazioniste dell'apprendimento proposte in ambito comportamentista possono essere classificate in base a
 - a) livello di analisi di S e R (*microscopico, macroscopico o molare*)
 - b) struttura delle associazioni (associazioni *dirette, mediate*)
 - c) legame tra S e R (*causale, non causale*)
 - d) meccanismo di modifica delle associazioni (p.es. *rinforzo, decadimento spontaneo, esercizio, interferenza, contiguità temporale*)
 - e) tipo di influenza del meccanismo di modifica delle associazioni (*deterministico, probabilistico*)

1) La teoria di **Thorndike**, fu sviluppata tra il 1898, anno di pubblicazione del suo lavoro fondamentale (*Animal intelligence, Psychological Review*) e il 1930. Egli si definì il fondatore del *connessionismo*, avendo introdotto per primo il concetto di *connessione tra stimolo e risposta*, ma oggi viene classificato come *funzionalista-analitico*. Sperimentò a fondo sugli animali, specialmente sui gatti nella sua **puzzle-box** (gabbia dei problemi) e i comportamentisti accolsero con favore le sue idee. E' una teoria macroscopica, di tipo diretto, causale, basata sul rinforzo, probabilistica. Secondo la **legge dell'effetto** la probabilità a priori, o "forza", della connessione è funzione del livello di soddisfazione/insoddisfazione prodotto dalla risposta emessa R (premio/punizione). Secondo Thorndike l'efficacia del rinforzo positivo è però maggiore rispetto a quella di un rinforzo negativo di pari intensità.

L'apprendimento è un processo per prove ed errori, non sostanzialmente influenzato dalle pulsioni interne (fisiologiche o cognitive). La pratica ha una certa importanza, avendo egli sostenuto la **legge dell'esercizio** (l'apprendimento migliora con la ripetizione delle prove). La teoria non giunse mai a una espressione quantitativa.

2) La teoria di **Edwin R. Guthrie** (1886-1959) è una teoria microscopica, di tipo diretto, causale, basata sulla contiguità temporale e interferenza, probabilistica. La natura della connessione S-R è la **contiguità temporale a livello microscopico**. I due principi della teoria affermano che

- a) se una particolare combinazione di stimoli S_1, \dots, S_n , è immediatamente seguita da una risposta R, allora è stabilita l'associazione $\{S_1, \dots, S_n\} \rightarrow R$
- b) la forza dell'associazione è massima alla prima occorrenza (**apprendimento tutto-o-nulla**)

In realtà, quello che Pavlov o Thorndike descrivevano con $S \rightarrow R$ è per Guthrie una sequenza $S \rightarrow R_0 \rightarrow S_1 \rightarrow R_1 \rightarrow S_2 \dots R_{n-1} \rightarrow S_n \rightarrow R$, dove i termini intermedi sono inosservabili in quanto microscopici. La risposta macroscopica R è predicibile solo in senso probabilistico. In tutto questo la pratica, le pulsioni e gli incentivi non hanno alcuna importanza e la **legge dell'effetto** di Thorndike è falsa. Il rinforzo positivo può servire unicamente a impedire per un certo tempo l'insorgere di

risposte diverse da quella premiata (basata sulla contiguità). L'estinzione non avviene per cessazione del rinforzo, ma per un'**interferenza** tra risposte diverse che possono risultare a partire da S.

3) La teoria di **Edward C. Tolman** (1886-1959), esposta nell'articolo "*Purposive behavior in animals and men*" (1932), è una teoria macroscopica, di tipo indiretto, causale, basata sull'aspettativa, probabilistica. Secondo Tolman il comportamento non è determinato da associazioni S-R, ma dagli scopi che si prefigge. Il comportamento è un seguire i **segni** che, dotati di un **significato** in relazione a una meta, la prefigurano e portano ad essa. L'associazione da stabilire non è dunque tra stimoli e risposte, ma tra segni e significati. La ripetuta esposizione a sequenze di segni che precedono una meta fa sì che aumenti la probabilità che ciascuno di questi segni venga interpretato come anticipatore della meta.

In un esperimento del 1930 tre gruppi di ratti affamati, introdotti all'ingresso di un labirinto, dovevano raggiungere un punto meta. I ratti del gruppo A trovavano sempre il cibo nel punto meta, quelli del gruppo C non lo trovavano mai, quelli del gruppo B lo trovavano solo a partire dal dodicesimo giorno di sperimentazione. Le prestazioni venivano misurate dal numero medio di errori (non raggiungimento del punto meta) compiuti in un giorno dai ratti di ciascun gruppo. Si osservò che per i primi 11 giorni le prestazioni del gruppo A erano superiori a quelle dei gruppi B e C, ma a partire dal dodicesimo giorno le prestazioni del gruppo B furono molto superiori a quelle del gruppo A. Un solo rinforzo aveva prodotto un crollo degli errori medi al di sotto di quelli compiuti dai ratti che avevano sempre ricevuto il rinforzo. Per Tolman i ratti del gruppo B avevano col tempo acquisito una **mappa cognitiva** del labirinto pur in assenza di rinforzi. Il concetto di mappa cognitiva rappresenta una rottura rispetto all'impostazione behavioristica tradizionale.

4) **Clark L. Hull** (1884-1952), nell'ambito del comportamentismo, ebbe una posizione opposta a quella di Skinner. La sua teoria macroscopica, di tipo indiretto, causale, basata sul rinforzo e sulle pulsioni interne, probabilistica, a più riprese rivista dalla prima versione del 1935 ("*The conflicting psychologies of learning – a way out*", Psychological Review) alla versione del 1952 (*A system of behavior*), ha una struttura matematica ipotetico-deduttiva basata su un certo numero di assiomi, e comprende una miriade di parametri legati a S, R, **rinforzo**, **generalizzazione**, **motivazione**, **inibizione**, **oscillazione**, **evocazione della risposta**. Il modello introduce da tre a cinque diversi livelli di variabili non osservabili tra S e R. I parametri di livello n determinano quelli di livello $n+1$. Parametri primari sono N , il numero dei rinforzi avuti fino al momento presente, C , la pulsione primaria (p.es. la fame, in ore di digiuno), S , l'intensità dello stimolo s , P , una misura del premio (p.es. quantità di cibo), W , lavoro necessario a emettere la risposta r .

Un parametro di secondo livello è sHr , o **forza dell'abitudine** (*habit*), la tendenza di uno stimolo s a produrre una risposta r .

Il **rinforzo primario**, è un aumento di sHr che si verifica sse:

- a) s è uno stimolo motivazionale legato a una pulsione primaria (p.es. la sensazione di fame)
- b) s e r sono temporalmente contigui
- c) r è associata a una rapida diminuzione di s .

Il **rinforzo secondario** si ha quando s è uno stimolo neutro, ma concomitante a uno primario.

sHr aumenta con il numero dei rinforzi N secondo la legge $sHr = 1 - 10^{-0.0305N}$.

Altri parametri di secondo livello sono V , legato all'intensità dello stimolo S dalla relazione $V = 1 - 10^{-0.44 \log S} = 1 - S^{0.44}$; D , l'intensità della pulsione (legata a C); K , la motivazione prodotta dall'incentivo, legata all'entità del premio P dalla relazione $K = 1 - 10^{-\alpha P}$; relativamente alla risposta r , la **forza dell'abitudine generalizzata** è $\sum s' Hr$, dove la sommatoria è estesa a tutti gli stimoli s' .

La grandezza più importante del terzo livello è il **potenziale di reazione** sEr , definito come

$$sEr = D \cdot V \cdot K \cdot sHr.$$

La natura moltiplicativa della relazione riflette l'importanza dei fattori di pulsione e motivazione data dall'incentivo oltre che della forza dell'associazione $s-r$. La formula fu in seguito corretta da **Spence** (1956) che sostituì $D \cdot K$ con $(D+K)$, per spiegare il presentarsi di risposte anche in assenza di premio ($K=0$).

Sempre al terzo livello, relativamente alla risposta r , Hull definisce il **potenziale di reazione generalizzato** come $\sum s' Er$, e un **potenziale inibitorio globale** $\sum s' Ir$, dove la sommatoria è estesa a tutti gli stimoli s' .

La probabilità di emissione della risposta r è qualcosa di simile a

$$Pr(r) = Pr[\sum (s' Er - s' Ir) > rth].$$

Dove rth è una **soglia di reazione**, e la sommatoria è il **potenziale netto di reazione** (sNr).

Al quinto livello, tra le grandezze che caratterizzano la risposta r , vi sono sTr , la *latenza* tra s e r ; A , l'*ampiezza* di r ; n , il *numero di emissioni* di r prima che si abbia l'estinzione in assenza di incentivi.

Si ha

$$sTr = (2.845/sNr)^{0.483}$$

La teoria di Hull è complessa e articolata al punto di assumere i caratteri della scienza cognitiva.

5) La **teoria della campionatura dello stimolo** (*Stimulus Sampling Theory, SST*) di **William K. Estes** (1950) è teoria microscopica, diretta, causale, basata sulla

contiguità temporale, probabilistica. Essa considera, come Guthrie, la formazione dell'associazione S-R un evento tutto o nulla. La presentazione di un particolare stimolo equivale per Estes all'estrazione casuale di un campione s da una popolazione degli elementi di stimolazione possibile S in un dato esperimento.

Ogni elemento s in $S = \{s_i\}$ è associato a una risposta r in $R = \{r_j\}$, e la probabilità di emissione della risposta r_j è uguale alla probabilità di estrazione di uno degli stimoli ad esso associati, cioè alla proporzione p_j degli stimoli s in S associati a quella risposta. Questa proporzione cambia in funzione degli eventuali rinforzi E che si fanno seguire a una data risposta. Se alla presentazione n -esima la risposta r_j viene fatta seguire da un rinforzo, la probabilità di emissione della stessa risposta alla prova seguente sarà

$$p_{n+1} = p_n + h(1 - p_n)$$

dove h è la probabilità di campionamento. Quindi

$$p_n = 1 - (1 - p_1)(1 - h)^{n-1}$$

Questa espressione descrive abbastanza bene la curva di apprendimento osservata in molti esperimenti. L'approccio di Estes si è però rivelato insufficiente per spiegare alcuni fenomeni come il recupero spontaneo e le curve di oblio. Estes rimediò in parte a queste insufficienze rendendo r dipendente dagli stati interni del soggetto. Una teoria analoga alla SST fu elaborata da **Voeks** nel 1950. Altre generalizzazioni hanno fatto uso della **Teoria dei Processi Markoviani**. Sulla base della teoria di Estes si è avuto lo sviluppo negli anni 60 di una serie di **Teorie Matematiche dell'Apprendimento**.

6) La teoria di **Rescorla e Wagner** (1972) si basa su un modello di apprendimento a variazioni continue e non tutto o nulla e riesce, a differenza della SST di Estes, a spiegare il fenomeno del **blocking**: inizialmente l'esperimento segue le regole del condizionamento classico, lo stimolo condizionato cs è seguito dallo stimolo non condizionato us . Nella seconda fase cs è presentato per un certo numero di volte insieme a uno stimolo neutro x , ed entrambi sono seguiti da us . Nelle prove di richiamo si osserva come la risposta condizionata non possa essere evocata dalla sola presentazione di x .

Questo fenomeno è in contraddizione con le previsioni di Estes. Rescorla e Wagner ipotizzarono un modello di apprendimento basato sulle formule

$$\begin{aligned} w_a^{n+1} &= w_a^n + F_a \cdot b \cdot (w_{max} - w_a^n - w_x^n) \\ w_x^{n+1} &= w_x^n + F_x \cdot b \cdot (w_{max} - w_a^n - w_x^n) \end{aligned}$$

dove w_a^n è la forza di associazione $cs \rightarrow cr$ alla prova n -esima, w_x^n è la forza di associazione $x \rightarrow cr$ alla prova n -esima, F_a l'intensità dello stimolo cs , F_x l'intensità dello stimolo x , b il *tasso di apprendimento*, w_{max} il valore dell'aspettativa.

Chiaramente l'apprendimento si verifica solo quando $w_a^n + w_x^n$ ha un valore lontano da quello di w_{max} , cioè quando gli eventi esterni sono “inattesi” rispetto alle nostre aspettative.

CIBERNETICA

- La **cibernetica**, teoria generale dei processi di controllo e di retroazione, fu inaugurata dal matematico americano **Norbert Wiener**. La parola, dal greco “*kybernétes*” (timoniere), era già stata usata da Platone e dal fisico francese Ampère per indicare l'arte del controllo politico-sociale. Allievo di D.Hilbert e B.Russell, Wiener fu professore al MIT dal 1932. Dal 1940 al 1942 lavorò per il governo alla realizzazione di un sistema di puntamento antiaereo. All'università di Città del Messico, con i fisiologi Cannon e Rosenblueth applicò gli stessi concetti di retroazione alla modellizzazione del comportamento del cuore. Del 1943 è l'articolo “*Behavior, Purpose and Teleology*”, scritto con **A.Rosenblueth** e **J.H.Bigelow**. Nel 1947 pubblicò “*Cibernetica, o controllo e comunicazione negli animali e nelle macchine*”. Successivamente molti autori portarono avanti la cibernetica, come Ashby, Von Neumann, McCulloch, Maturana, Pask, Grey Walter, Von Foerster.
- Nel 1943 **McCulloch e Pitts** pubblicarono “*A logical calculus of the ideas immanent in nervous activity*”, dove si propone un modello del funzionamento dei neuroni corticali basato su trasmissione dei segnali e retroazione,
- **Von Foerster**, fisico tedesco emigrato negli US, distinse tra **cibernetica di primo ordine** e **cibernetica di secondo ordine**. Quest'ultima non trascura la presenza dell'osservatore.

TEORIA DELL'INFORMAZIONE

- Nel 1948, con l'articolo “*A mathematical theory of communication*” di **Claude Shannon** e nel 1949, con il libro “*The mathematical theory of communication*” di **C.E.Shannon** e **W.Weaver**, fu fondata la **teoria dell'informazione**, secondo la quale l'informazione contenuta in un messaggio può essere quantificata indipendentemente dal sistema fisico di trasporto, e dal tipo di messaggio. Un canale di comunicazione può essere caratterizzato dalla quantità massima di informazione che vi può transitare. Il problema fondamentale della comunicazione è quello di riprodurre in maniera esatta o approssimata in un punto un messaggio selezionato in un altro punto. Normalmente i messaggi hanno un significato (*meaning*), ma ai nostri fini l'unico aspetto significativo è che il messaggio trasmesso ω è stato scelto da un insieme Ω .
- Gli elementi di un modello generale di processo di comunicazione prevedono
 - a) La **sorgente dell'informazione**, che produce un messaggio ω , o una sequenza di messaggi. Il messaggio può essere una sequenza di simboli, oppure una o più funzioni di una o più variabili

- b) Il **trasmettitore**, che codifica il messaggio in un segnale s adatto alla trasmissione sul canale (p.es. campionamento-compressione-quantizzazione-codifica in un sistema PCM, oppure FM)
- c) Il **canale di trasmissione**, il mezzo che trasporta il segnale (p.es. un cavo coassiale, una banda di frequenze radio, un fascio di luce)
- d) Il **ricevitore**, che opera sul segnale ricevuto la trasformazione inversa del trasmettitore
- e) La **destinazione**, la persona (o la cosa) cui il messaggio è destinato.

La teoria di Shannon si occupa solo della fase di trasmissione del segnale dal trasmettitore al ricevitore. Il generico segnale trasmesso s sarà uno tra N possibili segnali.

Shannon distingue tre tipi di sistemi di comunicazione:

- a) discreti (ω e s sequenza di simboli discreti)
 - b) continui (ω e s funzioni continue)
 - c) misti.
- Si ipotizza che il destinatario abbia delle *aspettative* e quindi un modello del comportamento del trasmettitore, espresso mediante una distribuzione di probabilità a priori $p(s)$. Quanto più il messaggio s ricevuto dal destinatario suscita sorpresa, in quanto $p(s) \ll 1$, tanta più informazione esso porta con sé.
 - Se N è finito allora N o una qualsivoglia funzione monotona di N ci dà una misura dell'informazione prodotta quando un messaggio è scelto in Ω , in modo equiprobabile. La scelta più conveniente è il logaritmo, in quanto
 1. parametri come il tempo, la larghezza di banda, ecc. tendono a variare col logaritmo delle possibilità. P.es. aggiungendo ad un sistema elettronico formato da K flip-flop un altro flip-flop, raddoppia il numero di stati del sistema, mentre aumenta di 1 il suo logaritmo in base 2.
 2. Il logaritmo semplifica molte espressioni matematiche.
- Scelto il logaritmo in base 2 l'unità di informazione è chiamata cifra binaria (*binary digit*), o **bit** come suggerì *J. W. Tukey*. Un dispositivo con due stati stabili può immagazzinare 1 *bit* di informazione. K dispositivi possono immagazzinare K *bits*, dato che il numero totale di stati possibili è 2^K , e $\log_2 2^K = K$.
- Def. Il **contenuto di informazione di un risultato** s è

$$h(s) = -\log_2 \Pr\{s\}$$

- In un sistema discreto, nel quale il trasmettitore possa trasmettere di volta in volta uno tra N segnali differenti $X = \{s_1, \dots, s_N\}$, con probabilità p_1, \dots, p_N , l'**informazione media** è

$$H = -\sum_{i=1..N} p_i \log_2 p_i$$

Per una v.a. quasi continua s , con densità di probabilità $p(s)$, avremo analogamente

$$H = - \int p(s) \log_2 p(s) ds$$

- Def. La **ridondanza di X** è

$$1 - [H/\log_2 N]$$

- La funzione di entropia soddisfa:

$$\begin{array}{ll} H \geq 0 & \text{dove l'uguaglianza vale sse } \exists i \mid p_i = 1 \\ H \leq \log_2 N & \text{dove l'uguaglianza vale sse } \forall i \mid p_i = 1/N \end{array}$$

- Th. $H \leq \log_2 N$ dove l'uguaglianza vale sse $\forall i \mid p_i = 1/N$
Dim. Vogliamo massimizzare la funzione di N variabili $\mathbf{p} \equiv (p_1, \dots, p_N)$

$$H(\mathbf{p}) = - \sum_{i=1..N} p_i \log p_i$$

soggetta al vincolo

$$\sum_{i=1..N} p_i = 1$$

Possiamo ricorrere al metodo dei **moltiplicatori di Lagrange**: all'equazione di vincolo associamo un moltiplicatore di Lagrange λ . Otteniamo l'equazione di Lagrange

$$F(\mathbf{p}) = - \sum_{i=1..N} p_i \log p_i + \lambda (\sum_{i=1..N} p_i - 1)$$

Derivando rispetto a ciascuna variabile p_i ,

$$\partial F / \partial p_i = - \log p_i - 1 + \lambda$$

In corrispondenza di un estremo per $H(\mathbf{p})$ le precedenti devono annullarsi, cioè

$$\log p_i = \lambda - 1$$

quindi le variabili p_i devono essere tutte uguali, sicché $p_i = 1/N$. Il fatto che questo estremo sia un massimo lo si deduce valutando la curvatura di $F(\mathbf{p})$, che risulta essere definita negativa, infatti

$$\partial^2 F / \partial p_i \partial p_j = - \delta_{ij} / p_i \leq 0$$

c.v.d

- Almeno dal punto di vista formale, H è esattamente l'entropia S così come introdotta da **Boltzmann** in meccanica statistica proporzionale al logaritmo del numero di stati possibili in cui può trovarsi il sistema (se i segnali sono equiprobabili $H = \log_2 N$). Nel caso dei sistemi termodinamici la costante di proporzionalità è la **costante di Boltzmann** k_B , e il logaritmo è naturale. Per quanto riguarda H la costante di proporzionalità non c'è e il logaritmo è in base 2, mentre l'unità di informazione è il *bit*.
- Le aspettative del destinatario sono soggettive, sicché destinatari diversi potrebbero associare allo stesso segnale valori diversi di informazione. L'alternativa è pensare di ricavare in modo oggettivo le probabilità di emissione p_i (p.es. identificandole con le frequenze relative osservate su un certo periodo di tempo, oppure mediante ragionamento). La questione della determinazione oggettiva delle probabilità a priori è però molto delicata e controversa.
- Il modo più semplice che il destinatario ha di stimare le probabilità di emissione alla sorgente è quello di contare le occorrenze dei vari messaggi. Distinguiamo il caso in cui il destinatario conosca il numero di messaggi diversi N , dal caso in cui non lo conosca.
- **N noto.** Inizialmente la stima è $p_i^1 = 1/N$ ($i=1, \dots, N$). Se al tempo $t_n > t_0$ il destinatario riceve il messaggio $\omega(t_n) \in \Omega = \{\omega_1, \dots, \omega_N\}$, il destinatario aggiorna le probabilità a priori come segue:

$$p_i^{n+1} = n p_i^n / (n+1) \quad \text{se } \omega(t_n) \neq \omega_i$$

$$p_i^{n+1} = (1 + n p_i^n) / (n+1) \quad \text{se } \omega(t_n) = \omega_i$$

L'informazione media cambia pure:

$$H^1 = -\sum_{i=1..N} p_i^1 \log_2 p_i^1 = \log_2 N$$

$$H^n = -\sum_{i=1..N} p_i^n \log_2 p_i^n$$

$$\begin{aligned} \Delta H^{n+1} &= -\sum_{i=1..N} p_i^n \log_2 p_i^n + \sum_{i=1..N} p_i^{n+1} \log_2 p_i^{n+1} \\ &= -\sum_{i=1..N} p_i^n \log_2 p_i^n + p_j^{n+1} \log_2 p_j^{n+1} + \sum_{i \neq j} p_i^{n+1} \log_2 p_i^{n+1} \\ &= -p_j^n \log_2 p_j^n + [1/(n+1) + n p_j^n / (n+1)] \log_2 [(1 + n p_j^n) / (n+1)] \\ &\quad + \sum_{i \neq j} p_i^n \{ [n/(n+1)] \log_2 [n/(n+1)] - [1/(n+1)] \log_2 p_i^n \} \end{aligned}$$

dove abbiamo indicato $\omega(t_n) = \omega_j$

- Shannon dimostrò anche che, fissate le caratteristiche del canale, la probabilità che non vi sia distorsione del segnale non può superare un limite massimo, e che è sempre possibile, almeno in teoria, raggiungere questo limite, pur di usare una opportuna forma di codifica.
- Def.

INTELLIGENZA ARTIFICIALE

- *Def.* L'**Intelligenza Artificiale (AI)** è la scienza che si occupa della progettazione di macchine...
 - Def.1* che pensano.
 - Def.2* capaci di esibire comportamenti intelligenti
 - Def.3* che esibiscono comportamenti che gli esseri umani attribuirebbero a processi mentali intelligenti
- Il problema di accertare le possibilità mentali di una macchina fu affrontato sin dai tempi dell'antica Grecia, ma solo dagli anni '50 divenne un problema non soltanto filosofico, di interesse per ricercatori, ingegneri, matematici, informatici, fisici.
- L'impostazione platonica, secondo la quale conoscenza e intelligenza sarebbero simboli e manipolazioni di simboli, ha ispirato anche il paradigma dell'*AI* e la prima Scienza Cognitiva.
- L'*AI* fu il primo tentativo di costruire una CS. Di fatto essa fu il primo nucleo della CS. Ciò consentì da un lato alla CS di usare concetti per costruire modelli e tecniche per verificarli sperimentalmente, ma ha anche distolto l'attenzione dai processi naturali negli uomini e negli animali e ha favorito l'atteggiamento riduzionistico, che oggi sta tramontando, secondo cui la teoria dei processi mentali sarebbe riducibile alle teorie computazionali.
- Tra i precursori dell'*AI* ricordiamo: Gottfried Wilhelm von Leibnitz, inventore della prima macchina calcolatrice; Charles Babbage e la contessa Ada Lovelace, creatori del primo computer in senso moderno; George Boole; Alan Mathison Turing; Hermann Hollerith; Vannevar Bush, che inventò il primo calcolatore analogico e incoraggiò il suo assistente C. Shannon a studiare nella sua tesi di dottorato la realizzabilità delle leggi Booleane mediante circuiti elettronici; John Von Neumann, che nel 1945 completò il primo calcolatore digitale, l'ENIAC.
- Nel 1954 si formò alla Carnegie Mellon University di Pittsburgh un gruppo guidato da **Allen Newell** e **Herbert Simon**. Nel 1956 in un seminario tenuto al Dartmouth College di Hannover, nel New Hampshire, dieci studiosi tra cui Chomsky, McCarthy, Miller, Minsky, Newell e Simon, proposero un progetto di ricerca su un ambito denominato "*Intelligenza Artificiale*". Nel 1957 Newell, Shaw e Simon realizzarono il programma **Logic Theorist**, capace di dimostrare teoremi di Logica. Il programma evolvette nel 1960 nel **General Problem Solver**, che avrebbe dovuto risolvere problemi utilizzando strategie generali indipendenti dal dominio specifico di interesse. Nel 1959 McCarthy proponeva la prima versione del linguaggio di programmazione LISP.
- Distinguiamo tre fasi nell'evoluzione storica dell'*AI* classica:
 - 1) Fase dei **metodi deboli**: si è convinti che esistano regole generali, indipendenti dal dominio del problema.

2) Fase dei **micromondi**: si cerca di caratterizzare la struttura dei domini specifici di conoscenza. L'ambito più studiato fu quello dei blocchi, di forma cubica o cilindrica, che il computer è in grado di muovere nello spazio. Del 1965 sono i programmi di Visione Artificiale di Roberts. SHRDLU di Winograd era invece in grado di comprendere semplici comandi in linguaggio naturale.

3) Fase dei **sistemi esperti**: si cerca di simulare valutazioni di un esperto umano in un determinato campo. Esempi sono DENDRAL, progettato dal 1965 da Feigenbaum e dedicato all'analisi delle strutture molecolari; MYCIN, progettato dal 1972 da Buchanan e Shortliffe e dedicato alla diagnosi delle malattie infettive; SHELL, capace di ragionare e trarre conclusioni sulla base delle conoscenze fornite dall'utente. Nel 1959 **Samuel** realizzò un programma per giocare a dama, **Genernter** scrisse un programma capace di dimostrare teoremi di geometria piana, mentre il programma di **Slagle** calcolava l'integrale indefinito di una funzione (1963).

- Def. Un sistema esperto è un programma progettato per risolvere problemi in un dominio specifico, con una efficienza paragonabile a quella di un esperto umano.
- I primi sistemi esperti erano composti da :
 - a) la **base di conoscenza**. Fatti e regole che connettono fatti, espressi in un linguaggio formalizzato (p.es. il Calcolo dei Predicati del Primo Ordine). Il lavoro è svolto da un ingegnere della conoscenza in base a un'indagine fatta su esperti umani.
 - b) Il **motore inferenziale**. Programma capace di estrapolare nuovi fatti o nuove regole dalla *base di conoscenza*.
 - c) L'**interfaccia utente**.
 - d) Il **modulo di aggiornamento della conoscenza**. Permette all'utente di introdurre nel sistema una nuova conoscenza.
 - e) Il **modulo di spiegazione**. Report per l'utente dei passi seguiti in un'inferenza
- I motori inferenziali sono costruiti sulla base di una delle seguenti regole:
 - 1) Regole di deduzione tradizionali.
 - 2) Regole di ragionamento probabilistico
 - 3) Regole euristiche.
 - 4) Regole di ragionamento basate sui casi.
- Tra i vantaggi di un sistema esperto rispetto a un esperto umano:
 - 1) Permanenza delle capacità e delle informazioni
 - 2) Replicabilità
 - 3) Basso costo
 - 4) Coerenza del comportamento
 - 5) Capacità di prendere in considerazione tutta la base di conoscenza
 - 6) Velocità
- Tra gli svantaggi:
 - 1) Incapacità di buon senso e creatività
 - 2) Difficoltà a manipolare ingressi sensoriali complessi

3) Incapacità di funzionare in presenza di ingressi incompleti, ambigui o mal definiti

- Il laboratorio di AI di Stanford fu fondato da **McCarthy**, già docente al MIT e collega di Minsky negli anni '50 (trasferitosi a Stanford nel 1962). A quel tempo McCarthy era già noto per aver inventato il LISP.

Al MIT si ebbe una maggiore sensibilità riguardo ai risultati ottenuti in campo psicologico e una maggiore disponibilità di finanziamenti.

- Negli anni '60 **Terry Winograd** lavorò al sistema **SHRDLU**, un programma con abilità cognitive nel **mondo dei blocchi**, uno spazio 3D con solidi triedrici (in ogni vertice si incontrano sempre 3 piani). SHRDLU non si occupava di riconoscimento percettivo ma di comprensione di comandi vocali relativi allo spostamento dei blocchi.
- Nel 1963 **Weizenbaum** costruì **ELIZA**, capace di emulare uno psicoterapeuta di scuola rogersiana (quindi con atteggiamento non direttivo). Il fatto che semplici regole sintattiche, assolutamente indipendenti dal contesto, producessero un effetto talmente realistico contribuì a suscitare consensi intorno alla CCS.
- Con gli studi di G.Roberts sulla Visione Artificiale (AV) questo divenne un campo di ricerca centrale al MIT.
- Nel 1964 Bertram Raphael realizzò SIR (Semantic Information Retrieval) capace di deduzioni logiche da una serie di premesse, mentre Daniel G.Bobrow costruì STUDENT, che risolveva problemi di algebra a livello di scuola media.
- Il gruppo di Stanford produsse negli anni successivi **DENDRAL**, iniziato nel 1965. DENDRAL, in base a dati sperimentali era in grado di avanzare ipotesi sulla formula di struttura dei composti chimici sconosciuti.

Nel 1972, sotto la guida di **Buchanan e Shortliffe**, fu scritto **MYCIN**, che dava diagnosi mediche sulle malattie ematiche in base ai dati di anamnesi e analisi.

Il motore inferenziale e l'interfaccia utente di MYCIN furono usati per costruire EMYCIN, la prima versione di SHELL (la base di dati deve essere fornita dall'utente). Seguirono una serie di sistemi esperti in campo medico, PUFF, CENTAUR, CASNET, ecc.

- Sempre al MIT furono conati due metodi fondamentali di rappresentazione della conoscenza: i frames (Minsky, 1975) e gli scripts (Schank, 1977).
- I sistemi esperti hanno avuto una certa fortuna dal 1980 al 1987, quando si stimò che ci fossero 1900 sistemi esperti impiegati in attività produttive, per un giro d'affari di 1 miliardo di dollari. All'inizio degli anni '90 si è avuta una certa ripresa con l'avvento delle **reti neurali** (NN) e della **logica fuzzy** (FL).
- I tre principi fondamentali che contraddistinsero il lavoro di **Allen Newell** e **Herbert Simon** nella seconda metà degli anni 50 furono:

a) Definizione comportamentale di intelligenza.

Def. Un programma manifesta **intelligenza** quando origina comportamenti che gli esseri umani giudicherebbero intelligenti.

Questa definizione è ispirata al behaviorismo e alla teoria della computazione di Turing. Il **test di Turing**. Nel 1950 Turing scrisse l'articolo "*Computing Machinery and Intelligence*", dove immaginava di poter operazionalizzare il possesso di stati di consapevolezza da parte di una macchina. C'è un uomo che ha il ruolo di interrogante (A), un altro uomo (B) e la macchina, un computer su cui gira un apposito progeamma (M). A non può vedere né sentire B e M, dal momento che si trova in una stanza separata, ma può loro rivolgere domande e ottenere risposte per mezzo di una tastiera e un monitor. A non sa a priori chi dei due è l'uomo e chi la macchina. Il suo compito è appunto stabilirlo sulla base delle risposte ottenute. Scopo di B è convincere A di essere l'uomo; scopo del programma M indurre A a sbagliare. Secondo Turing M può essere ritenuto intelligente se induce A a dare una valutazione errata.

Il punto di vista di Turing è quello dei sostenitori dell'AI forte, che fanno risalire la data di nascita dell'AI proprio al 1950, anno di pubblicazione dell'articolo di Turing.

b) Ipotesi del **Sistema fisico di simboli** (*Physical Symbol System*):

- l'intelligenza è possibile grazie all'esistenza e alla manipolazione di rappresentazioni simboliche di eventi e situazioni sia interni che esterni
- esiste un sottoinsieme minimale di regole di base, condiviso da tutti i sistemi cognitivi, che guida la manipolazione di questi simboli
- un modello di sistema cognitivo umano può essere sempre espresso sotto forma di un programma che, quando gira, produce il comportamento cognitivo effettivamente osservato negli esseri umani

c) Introduzione di **regole euristiche**. A differenza degli algoritmi, che producono risposte necessarie in un numero definito di passi, ma che spesso non sono noti o non sono praticamente applicabili, le *regole euristiche* sono strategie generali, estrapolate dall'esperienza accumulata dagli esseri umani su problemi simili, che guidano il programma verso la soluzione. Un esempio sono i problemi di **criptoaritmetica**.

- La **ricerca euristica** è l'applicazione di regole euristiche che selezionano solo un sottoinsieme delle possibili ramificazioni ad albero in cui si espande lo spazio del problema, e precisamente quelle più pertinenti e "promettenti" in base alle informazioni raccolte. L'obiettivo della ricerca euristica è quello di evitare l'esplorazione esaustiva di tutte le ramificazioni generabili.
- Nel 1957 Newell, Shaw e Simon realizzarono il programma **Logic Theorist**, capace di dimostrare teoremi di Logica Matematica. In questo dominio l'applicazione di regole euristiche è indispensabile per decidere quali regole di deduzione, e in che ordine, devono essere applicate per arrivare a dimostrare un teorema. Le regole euristiche usate nel Logic Theorist erano del tipo

se [condizione] allora [azione]

Inoltre era presente un interprete capace di tradurre uno stato generico del programma (assiomi e teoremi già dimostrati) nella *[condizione]* di qualche regola euristica e, a parità di applicabilità, decidere quale regola mettere in atto.

Il Logic Theorist fu applicato ai primi 52 teoremi del secondo capitolo dei Principia Mathematica di Russell e Whitehead, e riuscì a dimostrarne 38. Una delle dimostrazioni fu giudicata più elegante di quella presente nel libro.

- Il programma evolvette nel 1960 nel **General Problem Solver (GPS)**, che avrebbe dovuto risolvere problemi utilizzando regole euristiche indipendenti dal dominio di interesse. Per decidere quali regole introdurre nel sistema Newell e Simon osservarono soggetti umani impegnati in problemi diversi. Il GPS vantava dei miglioramenti rispetto al *Logic Theorist*, ed era incentrato su una **analisi mezzi-fini**. Si introdusse un modo per misurare la distanza tra lo *stato attuale* e lo *stato obiettivo* (la soluzione del problema o di un sottoproblema), e si cercava una strategia tra quelle disponibili (i “mezzi”) per minimizzare questa distanza e giungere allo *stato-meta* della soluzione (il “fine”). Il limite del GPS era la sua inapplicabilità ai problemi non formalizzabili simbolicamente (come p.es. quelli di criptoaritmetica, logica e scacchi) e a quelli senza soluzione nota a priori.
- La prima grossa difficoltà fu la traduzione automatica di un testo da una lingua all'altra, operazione che implica aspetti semantici di dipendenza dal contesto, la nozione delle conseguenze di un'azione e la memoria episodica. Un'altra difficoltà fu l'irriducibilità di un problema a un altro, che infranse l'illusione di un *General Problem Solver*. In seguito a queste difficoltà si optò per una divisione della conoscenza in **c. dichiarativa** (sistema chiuso, che non produce effetti) e **c. procedurale** (produce effetti, diretta a uno scopo, dipende dal contesto).
- Riguardo all'AI, **D.R. Hofstadter** riporta il cosiddetto **teorema di Tesler**: “L'AI è tutto ciò che non è ancora stato fatto”. Questo enunciato fu suggerito dal fatto che, storicamente, non appena si è riusciti a programmare qualche funzione o abilità ritenuta mentale, subito si è smesso di considerarla come un ingrediente essenziale del “vero pensiero”.
- Nella CS la contrapposizione tra impostazioni platonica e aristotelica, o tra conoscenze dichiarativa e procedurale, rivive negli approcci computazionali *simbolico* e *subsimbolico*.

Secondo l'**approccio computazionale simbolico**, i processi cognitivi sono manipolazioni di simboli secondo regole codificabili in un computer. La potenza della mente è sostanzialmente dovuta non tanto alla complessità degli algoritmi, quanto alle sue enormi capacità di memoria. Il punto chiave è la definizione dei simboli, associati a proprietà elementari dei fenomeni.

Punti di forza sono:

- a) **iteratività** (connessione di sequenze simboliche)
- b) **ricorsività** (riporta l'uscita di una funzione al suo ingresso)

c) **ereditarietà** (una struttura simbolica subordinata a una più generale ne acquisisce automaticamente le proprietà)

d) **astrazione** (inferenza di una struttura simbolica sovraordinata)

I limiti sono:

a) necessità di molta memoria

b) sensibilità ai disturbi, alla incompletezza dei dati, al malfunzionamento delle unità

c) i simboli e le loro relazioni non hanno significato in sé, ma solo per noi

d) previsione e controllo limitati al crescere della complessità

L'approccio platonico-simbolico, capace di descrivere la struttura di una conoscenza di un dato soggetto in un dato momento, è però incapace di spiegare come si è arrivati a questo stato di cose e come l'interazione organismo-ambiente lo modificherà.

Secondo l'**approccio subsimbolico** i processi cognitivi sono fenomeni macroscopici emergenti dall'azione cooperativa di un gran numero di unità microscopiche, prive di significato molare. L'applicazione consistente con tale definizione sono le **reti neurali (NN)**, introdotte da **McCulloch e Pitts** (1943).

Punti di forza sono:

e) Possibile interpretazione neurofisiologica

f) Semplicità di definizione

g) Possibilità di elaborazione parallela

e) Immunità ai disturbi, al rumore, alla incompletezza dei dati

I limiti sono:

f) Mancanza di iteratività, ricorsività, ereditarietà.

g) i livelli di attivazione delle unità non hanno significato in sé, ma solo per noi

h) incapacità di recupero spontaneo di un'informazione, senza uno stimolo di richiamo

- Due approcci al problema della definizione del **significato** di un'espressione simbolica: l'approccio **strutturale** (il significato risiede nelle relazioni tra espressioni, p.es. le **reti semantiche**) e l'approccio **referenziale** (il significato risiede nelle relazioni con enti o eventi, i **referenti** dell'espressione). Dati i limiti delle reti semantiche ci si è sempre più orientati verso l'approccio referenziale.
- Negli anni '30 il logico polacco **Alfred Tarski** definì l'interpretazione semantica in una teoria logico-formale referenziale (o **estensionale**), espressa unicamente con:
 - a) **costanti individuali** (p.es. "π", "32", "pippo")
 - b) **costanti predicative** (p.es. "è un numero pari", "maggiore di")
 - c) **variabili individuali** (p.es. x , y)
 - d) **variabili predicative** (p.es. P , Q)
 - e) **connettivi logici** (p.es. "e", "oppure")
 - f) **quantificatori** (p.es. "∀", "∃")

g) *simboli ausiliari* (p.es. “(”,”), “=”, “+”)

Tarski ipotizza l'esistenza di **Universi**, insiemi di elementi, astratti (come i numeri naturali \mathbb{N}) o concreti. Scelto un Universo occorre definire una legge di corrispondenza biunivoca tra costanti individuali e elementi dell'Universo, tra costanti predicative binarie, ternarie,... e coppie, triple,... di elementi dell'Universo. Il significato delle **proposizioni** (combinazione di costanti individuali e predicative, p.es. “32 è maggiore di 23”) coincide col **valore di verità** (*vero*, *falso*) ad esse associato nel particolare Universo in cui sono interpretate. Tarski ipotizzava di poter sempre verificare con un numero finito di operazioni se un dato elemento appartiene o no a un determinato sottoinsieme dell'Universo. Oggi sappiamo che tale postulato è una pretesa infondata, anche sul piano puramente logico. L'approccio di Tarski fu ripreso negli anni '60 da **Kripke**, che generalizza la nozione di Universo con quella di **mondi possibili**, e fa uso di **operatori modali** (p.es “è necessario che...”, “è possibile che...”).

- L'impossibilità di tradurre la comprensione del significato in una semplice manipolazione di simboli fu enfatizzata dal filosofo **John Searle** con il suo celebre **argomento della stanza cinese**. Egli immagina un uomo, che non conosca affatto la lingua cinese, chiuso in una stanza, impossibilitato a comunicare con l'esterno se non per mezzo di una fessura, attraverso la quale possono passare solo tavolette. Di quando in quando l'uomo riceve una tavoletta con un ideogramma cinese, che sta per una domanda. Egli consulta un catalogo in cui a ogni possibile ideogramma ne è associato uno di risposta, quindi emettere all'esterno una tavoletta con l'ideogramma di risposta. All'esterno della stanza ci sono veri cinesi che pongono domande in cinese e ottengono risposte corrette, in cinese. L'uomo manipola simboli, ciò nondimeno non capisce nulla di cinese. Il significato non ha a che vedere con la manipolazione di simboli, ma con la nostra struttura biologica, il nostro vivere in un tempo e in un luogo precisi, ecc.
- L'argomento della stanza cinese ha avuto molte repliche. La più interessante è quella di Fodor ed è chiamata la “**replica del robot**”. Se un robot eseguisse la manipolazione delle tavolette per un tempo sufficientemente lungo e avesse sensori che gli permettessero di associare gli ideogrammi che riceve al comportamento dei cinesi all'esterno della stanza certamente imparerebbe il cinese.
- In effetti Searle pensava a programmi statici, non a programmi capaci di modificarsi con l'esperienza. Questa evoluzione guidata dall'input ambientale potrebbe non essere descrivibile da una TM, e una volta raggiunta una qualche complessità critica mostrare una qualità emergente simile alla coscienza.

CONNESSIONISMO

- Nel 1949 lo psicologo behaviorista canadese **Donald Hebb** formulò la **teoria delle assemblee cellulari** (*The Organization of Behavior*, 1949). La teoria ricercava il substrato neurofisiologico della "circolazione interna" dell'informazione, sotto forma dei **circuiti riverberanti** individuati nel cervello già dagli anni 30 da neuroanatomisti come *Lorente de No*. Hebb vedeva i neuroni come semplici dispositivi ingresso-uscita che si attivano se la somma dei segnali in ingresso supera una certa soglia. L'**assemblea cellulare** è un insieme di neuroni con connessioni reciproche ad elevata efficacia sinaptica, sicché l'eccitazione di uno di essi provoca con alta probabilità l'eccitazione di tutti gli altri. Tale eccitazione collettiva può restare confinata all'assemblea per la presenza di circuiti riverberanti, oppure propagarsi all'esterno verso altre assemblee. L'esistenza di circuiti riverberanti permetteva di spiegare il differimento temporale tra stimolo e risposta (latenza). La rappresentazione mentale di un oggetto percepito sarebbe l'eccitazione di una o più assemblee in cui ciascun neurone codifica una particolare proprietà dell'oggetto (colore, forma, ecc.). La posizione di Hebb sulla formazione delle assemblee cellulari era rigidamente empirista: egli formulò l'ipotesi della **facilitazione sinaptica** (o **legge di Hebb**): ogniquale volta l'eccitazione di un neurone A è immediatamente seguita (dopo un tempo pari al ritardo sinaptico) dall'eccitazione di un neurone B, aumentava l'efficacia sinaptica della connessione tra l'uscita di A (l'assone) e l'ingresso di B (dendriti). Secondo Hebb la facilitazione sinaptica richiede un periodo di consolidamento che va da 15 min a 1h. Oggi esiste un gran numero di formulazioni quantitative differenti della legge di Hebb. Questa legge è stata in parte confermata negli anni '70 dagli studi di Bliss, Gardner-Medwin e Lomo. Essi descrissero un fenomeno di **potenziamento a lungo termine** nell'area dentata della formazione ippocampale del ratto. Registrando la risposta dei neuroni a seguito della stimolazione controllata di altri neuroni essi osservarono un effetto di sensibilizzazione, con una componente transitoria che si estingueva dopo la cessazione della stimolazione, e una componente a lungo termine.
- Il **Connessionismo** ha inizio nel 1943 con la **teoria delle reti nervose formali** di **McCulloch e Pitts**, modelli astratti di funzionamento della corteccia cerebrale. Nel 1956 **Rochester, Holland, Haibt e Duda** effettuano la prima simulazione e trovano le condizioni che garantiscono la formazione delle assemblee cellulari descritte da Hebb. Nel 1958 **Frank Rosenblatt** introduce il **Perceptrone**, capace di apprendere dall'esperienza e classificare patterns di ingresso. Nel 1969 **Minsky e Papert** dimostrarono che il Perceptrone è incapace di risolvere semplici problemi di classificazione, come distinguere una figura connessa da una non connessa. **Grossberg e Kohonen** elaborano nuovi schemi di apprendimento che non sfruttano la conoscenza delle risposte corrette. Nel 1982 **Hopfield** propone un modello di memoria associativa, in cui l'informazione richiamata dalla memoria dipende unicamente dalla somiglianza con uno stimolo di richiamo. Egli mostra

anche che le NN non sono altro che modelli di comportamento dei *vetri di spin*, speciali sistemi già noti nella fisica degli stati condensati. Il connessionismo si afferma pienamente a partire dal 1985 quando **Rumelhart, Hinton e Williams** propongono la regola di apprendimento *error-backpropagation* per il **Perceptrone multistrato**, capace di ovviare alle critiche mosse al vecchio Perceptrone. Nel 1989 **Hornik, Stinchcombe e White** dimostrano matematicamente che in effetti il Perceptrone multistrato è, in linea di principio, capace di apprendere qualunque relazione di ingresso-uscita. Nel 1986 **McClelland e Rumelhart** pubblicano l'antologia ***Parallel Distributed Processing. Exploration in the Microstructure of Cognition***.

TEORIA DEI LINGUAGGI FORMALI

- La teoria dei linguaggi formali è uno dei settori più attivi nell'ambito della scienza dell'informazione. Molti risultati teorici hanno contribuito enormemente allo sviluppo dei linguaggi di programmazione, alla comprensione dei linguaggi naturali, alla teoria della complessità, ecc.
- Secondo il paradigma di Turing i processi di calcolo, o algoritmi, devono avere un carattere discreto e deterministico. Non tutti i processi fisici sono computabili. P.es. non sono computabili il moto di un pianeta, in quanto non discreto, o il decadimento radioattivo, in quanto casuale.
- La parola **algoritmo** deriva dal nome del matematico persiano del IX secolo *Abu Gia'far Muhammad ibn Musa al-khwarazmi*, autore del testo matematico *Kitab al-Giabr wa'l-muqabalah*. La forma latina medievale del suo nome era **Algorismus**. Da *al-giabr* ("trasporto") deriva invece "algebra".
- *Def.* Un algoritmo è un insieme finito di istruzioni per dividere i numeri naturali in due classi, che dà sempre una risposta, in un tempo finito, e sempre la stessa per ogni numero fissato.
- Nella **Teoria dei Linguaggi Formali** si studiano sistemi e macchine idealizzati. Infatti non compare mai un riferimento al tempo, né allo spazio.

Definizioni Generali

- *Def.* Un **alfabeto** Σ è un insieme finito e non vuoto di **simboli**.
- *Def.* Una **stringa** è una sequenza finita di simboli appartenenti a Σ
- *Def.* La **lunghezza** $|w|$ di una stringa w è il numero dei simboli che la compongono.
- *Def.* La **stringa vuota** (o **simbolo nullo**) è la stringa di lunghezza 0. Per comodità la indichiamo con ε .
- *Def.* l'operazione binaria non commutativa e associativa di **concatenazione** (o **giustapposizione**) di due stringhe ordinate (w, v) è la stringa z ottenuta scrivendo le due stringhe una di seguito all'altra, $z = wv$. L'operazione estesa agli insiemi di

stringhe è $WV = \{wv \mid w \in W, v \in V\}$. L'elemento neutro della concatenazione è ε :
 $\forall w \in W \quad \varepsilon w = w\varepsilon = w$.

Def. $\Sigma^{n+1} = \Sigma^n \Sigma$; $\Sigma^0 = \{\varepsilon\}$

Def. $\Sigma^* = \{\varepsilon\} \cup \Sigma \cup \Sigma^2 \cup \Sigma^3 \dots$ (**operazione star di Kleene**)

Def. $\Sigma^+ = \Sigma^* - \{\varepsilon\} = \Sigma \cup \Sigma^2 \cup \Sigma^3 \dots$

Def. $\forall w, v \in \Sigma^*$: $w+v = (w \text{ oppure } v)$ (**disgiunzione, o somma logica**)

- Def. Σ^+ e l'operazione di concatenazione formano una struttura algebrica detta **semigruppso libero generato da Σ** .
- Def. Σ^* e l'operazione di concatenazione formano una struttura algebrica detta **monoide libero generato da Σ** .
- Def. **stringa riflessa** di $w = ad\dots ec \in \Sigma^*$ è $w^R = ce\dots da$
- Def. Una stringa w è palindroma $\Leftrightarrow w = w^R$
- Def. $n_a(w)$ denota il numero di occorrenze del simbolo $a \in \Sigma$ in $w \in \Sigma^*$

Linguaggio

- Dato un alfabeto di **simboli terminali** $\Sigma_T = \{a, b, \dots, c\}$, e l'insieme di tutte le concatenazioni possibili dei suoi elementi $\Sigma_T^* = \{\varepsilon, a, ab, aab, \dots\}$, un **linguaggio formale** su Σ_T è un sottoinsieme $L \subseteq \Sigma_T^*$. Le stringhe in L sono dette **sequenze** (o **parole**)
- Una **presentazione** di un linguaggio L è un modo di specificare quali sequenze gli appartengono. Alla semplice enumerazione, impraticabile per $|L|$ elevato, è preferibile la specificazione di **regole di produzione**, stringhe di segni (**metalinguistici**) definite da un **metalinguaggio**, applicando ripetutamente le quali si ottengono tutte e sole le sequenze di L .
- L è un insieme, quindi tutte le operazioni standard tra insiemi si applicano ai linguaggi producendo altri linguaggi.

Grammatica

- Una **grammatica** G è un insieme di **regole di produzione** del tipo $\alpha \rightarrow \beta$. Di G può darsi una **descrizione concreta** oppure una **descrizione schematica**. Quest'ultima usa **sequenze generiche** (o **simboli non terminali, variabili**) $\Sigma_N = \{S, A, B, \dots\}$. Uno dei simboli in Σ_N è scelto come **simbolo iniziale** (sia S). Definito allora l'insieme delle produzioni $P = \{\alpha \rightarrow \beta \mid \alpha \in \Sigma^+, \beta \in \Sigma^*\}$, la grammatica è una quadrupla

$$G = \langle \Sigma_T, \Sigma_N, P, S \rangle$$

dove si è definito $\Sigma \equiv \Sigma_T \cup \Sigma_N$.

- Abbiamo detto che le produzioni sono del tipo $\alpha \rightarrow \beta$. Ponendo delle restrizioni sulla struttura di α , di β o di entrambi, otterremo classi differenti di grammatiche.

- Il meccanismo generativo di una grammatica G si basa su catene di derivazione della forma $S \Rightarrow w_1 \Rightarrow w_2 \Rightarrow \dots \Rightarrow w_s$
dove $w \Rightarrow w'$ sse in w esiste una sottostringa che sia il primo membro di una produzione $\alpha \rightarrow \beta$ e tale che w' si ottenga da w per sostituzione di α con β . Le w_i sono dette **forme sentenziali** di G . Diremo che la derivazione genera w_s se $w_s \in \Sigma_T^*$.
- *Esempio1*: Il linguaggio $L = \{a, aa, aaa, \dots\}$, anche scritto $L = \{a^n | n > 0\}$, può essere generato dalla grammatica (tralasciando di specificare ogni volta Σ_T e Σ_N)

$$S \rightarrow a$$

$$S \rightarrow aS$$

- *Esempio2*: Il linguaggio $L = \{ab, aabb, aaabbb, \dots\}$, anche scritto $\{a^n b^n | n > 0\}$ può essere generato dalla grammatica

$$S \rightarrow ab$$

$$S \rightarrow aSb$$

- Fondamentalmente una grammatica funge da filtro: da tutte le sequenze possibili presenti in Σ_T^* estrae soltanto quelle significative
- *Def.* Sia $L(G)$ il linguaggio su Σ_T formato da tutte e sole le sequenze generate da G .

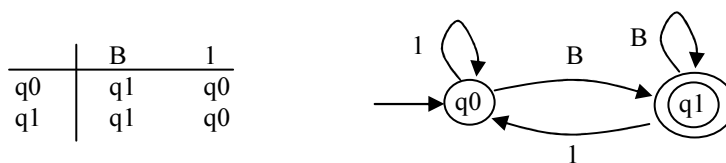
Automa

- Un **automa** (*automaton*, A) è un modello di macchina di calcolo tempo-discreta a stati finiti matematicamente idealizzata, nel senso che funziona per un tempo lungo a piacere senza mai commettere errori. Ha una qualche forma di *input* e di *output*, può avere o meno una qualche memoria di lavoro. L'input e l'output sono per definizione quantizzati. P.es. se l'input fosse una temperatura, l' A potrebbe acquisire solamente i valori nell'intervallo 0° - 100° con incrementi di 0.1°
- L'astrazione dai dettagli costruttivi della macchina è spinta fino al punto che il residuo di questa operazione è un insieme finito di **stati interni dell' A** , distinguibili ma fisicamente indefiniti. L' A acquisisce l'input sotto forma di stringhe su un supporto sequenziale di informazioni (*nastro di input*). Il nastro è dato come una sequenza di caselle in ciascuna delle quali è registrato un simbolo, scelto da un *alfabeto di ingresso* Σ_T , oppure nessun simbolo (" $\#$ ", " B ", *blank*). Nel caso più semplice il simbolo è uno solo, p.es. " I ".
- *Def.* Un A con output booleano è detto **accettore** (*Acceptor*) o **riconoscitore**.
- Un accettore A è usato per definire il *criterio di appartenenza* di una stringa a un linguaggio.
- *Def.* Sia $L(A)$ il linguaggio su Σ_T formato da tutte e sole le stringhe accettate da A .

- Def. Due automi accettori A_1 e A_2 sono **equivalenti** $\Leftrightarrow L(A_1) = L(A_2)$

DFA

- Il tipo più semplice di A è l'**automa finito deterministico (DFA)**, o semplicemente **FA**. Esso ha un numero finito di stati, $Q = \{q_0, \dots, q_{|Q|-1}\}$. Possiamo assumere che q_0 rappresenti lo stato iniziale. In generale potrebbero esserci più stati finali $q_j \in F \subseteq Q$. Il DFA è caratterizzato dalla **funzione di transizione** $\delta: Q \times \Sigma_T \rightarrow Q$, che stabilisce in modo deterministico (senza possibilità di scelta) lo stato al tempo $n+1$ noto l'ingresso e lo stato al tempo n . La δ può essere espressa in forma tabulare oppure come grafo



Il DFA è dotato di un nastro di input che scorre sequenzialmente in una sola direzione (sinistra) sotto una testina di sola lettura. Il nastro contiene inizialmente una sequenza di lunghezza finita $w \in \Sigma_T^+$, e la testina di lettura è posizionata sul primo simbolo all'estrema sinistra. A ogni passo il DFA rileverà presenza o assenza del simbolo e, in caso di presenza:

- 1) determinerà lo stato successivo in base alla funzione δ
- 2) determinerà lo scorrimento del nastro a sinistra di una posizione.

La lettura del nastro procede di pari passo con l'evoluzione temporale del DFA, che quindi terminerà esattamente dopo $|w|$ passi.

- Def. $w \in L(DFA) \Leftrightarrow$ lo stato raggiunto dall'automa dopo $|w|$ transizioni è finale.
- In definitiva un DFA è una quintupla

$$DFA = \langle \Sigma_T, Q, \delta, q_0, F \rangle$$

- Algoritmo di un DFA:

```

q = q0;
s = readTape();      /* legge il primo simbolo di w a sinistra */
while (s ∈ Σ_T) do
{
    q = δ(q,s);
    moveTapeLeft();
    s = readTape();
}
if q ∈ F then w ∈ L(DFA) else w ∉ L(DFA);

```

- Il *DFA* descritto è noto come **accettore a stati finiti di Rabin e Scott**
- Volendo dare una misura della complessità dei DFA che possono essere definiti su Q e Σ_T dati, sembra appropriato associarla al numero delle possibili funzioni di transizione. La funzione δ deve specificare per ogni stato ($|Q|$ in tutto) e per ogni simbolo in ingresso ($|\Sigma_T|$ in tutto) uno stato tra $|Q|$ possibili, quindi il numero di possibili funzioni di transizione è $|Q|^{|Q||\Sigma_T|}$. Notiamo che questo è un modo del tutto generale di procedere: dato un insieme di applicazioni $\Delta \equiv \{f: A \rightarrow B\}$ vale $|\Delta| = |B|^{|A|}$.
- Possiamo definire anche una **funzione di transizione estesa** $\delta^*: Q \times \Sigma_T^* \rightarrow Q$, sicché

$$L(DFA) = \{w \in \Sigma_T^* \mid \delta^*(q_0, w) \in F\}$$

- Un *DFA*, oltre che riconoscere le sequenze di un linguaggio può anche generarle (purché la testina abbia capacità di scrittura). Si può dimostrare che le due capacità si implicano a vicenda, almeno se ci si riferisce ad automi di una stessa categoria
- Un automa può anche essere usato come trasduttore (*transducer*) di sequenze in ingresso in sequenze di uscita.
- *Def.* Due **automi trasduttori** A_1 e A_2 sono **equivalenti** $\Leftrightarrow A_1$ e A_2 computano la stessa funzione

NFA

- Un *FA non-deterministico (NFA)* può essere tale in due sensi:
 - a) Uno stato può avere due o più archi che partono da esso labellati con lo stesso simbolo. Quando il simbolo è letto in ingresso, entrambe gli archi possono essere seguiti;
 - b) Uno stato può avere un arco che parte da esso labellato con ε . Questa transizione può avvenire senza che si abbia la lettura di un simbolo.

La funzione di transizione di un *NFA* è $\delta: Q \times (\Sigma_T \cup \{\varepsilon\}) \rightarrow 2^Q$, dove 2^Q denota il **powerset** di Q , cioè l'insieme di tutti i suoi sottoinsiemi.

Un accettore *NFA* accetterà la stringa w se almeno uno degli stati raggiunti dopo $|w|$ passi è finale, cioè

$$L(NFA) = \{w \in \Sigma_T^* \mid \delta^*(q_0, w) \cap F \neq \emptyset\}$$

- Un'altra generalizzazione dei *DFA* si ha permettendo transizioni di stato senza movimento del nastro. Come i *NFA*, anche queste macchine sono in grado di accettare solo *linguaggi regolari*.
- Può aver senso determinare il riconoscitore di un dato linguaggio avente minimo numero di stati. In genere questo sarà un *NFA*.

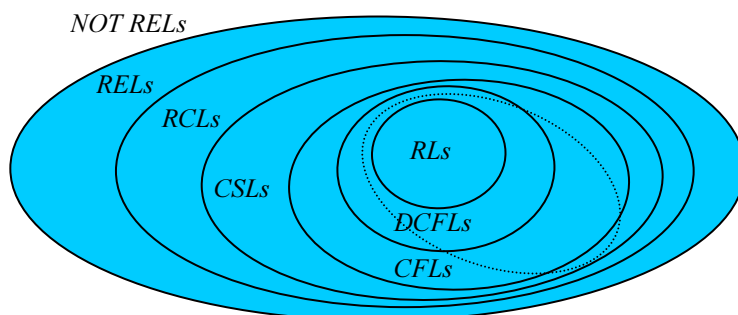
- I *FAs* non hanno memoria, o meglio la loro memoria è un unico registro, quello che tiene traccia dello stato corrente q . Il nastro può essere solamente letto, o al più scritto per emettere un output, non un risultato parziale da usare in seguito.

Gerarchie di linguaggi

- Noam Chomsky dimostrò che i *FAs* non possono simulare qualunque grammatica o accettare qualunque linguaggio. Egli individuò una gerarchia di linguaggi e di corrispondenti grammatiche e automi (**gerarchia di Chomsky**)

Livello	Linguaggio	Grammatiche	Automi
0	L. ricorsivamente enumerabili (REL) <i>Funzioni computabili</i>	G. non ristrette (UG)	Turing Machines (TM)
1	L. dipendenti dal contesto (CSL) <i>Esempio: $\{a^n b^n c^n\}$</i>	G. dipendenti dal contesto (CSG)	A. Lineari limitati (LBA)
2	L. indipendenti dal contesto (CFL) <i>Esempio: $\{a^n b^n\}$</i>	G. indipendenti dal contesto (CFG)	A. Pushdown non deterministici (NPDA)
3	L. regolari (RL) <i>Esempio: $\{a^n\}$</i>	G. regolari (RG) - G. lineari a destra (RLG) - G. lineari a sinistra (LLG)	DFA a NFA

- Nella gerarchia di Chomsky gli elementi di un dato livello includono quelli di livello superiore, sicché $\{REL\} \supset \{CSL\} \supset \{CFL\} \supset \{RL\}$.
- Oggi sappiamo che esistono decine di classi di linguaggi che la gerarchia di Chomsky non contempla. P.es. I *CFLs* possono essere suddivisi in **deterministici (DCFL)** e **non deterministici (CFLs)**, con $\{CFL\} \supset \{DCFL\}$; I rispettivi riconoscitori sono gli **automi pushdown deterministici (DPDA)** e **non deterministici (NPDA)**. E ancora, i **linguaggi lineari (LL)** si collocano tra i *RLs* e i *CFLs*, ma esistono *LLs* che non sono *CFLs* e *CFLs* che non sono *LLs*. Ancora, la classe dei **linguaggi ricorsivi (RCL)**, linguaggi per i quali esiste una TM che può sempre decidere per l'appartenenza di una stringa $w \in \Sigma_T^*$, include $\{CSL\}$, ma è inclusa da $\{REL\}$.
- Il fatto che $\{REL\} \supset \{RCL\}$ esprime in forma astratta il *Teorema di Gödel* che stabilisce che ogni sistema assiomatico consistente e adeguato a esprimere un insieme di asserzioni sufficientemente ampio sui numeri interi è incompleto, cioè possiede teoremi veri che non possono essere provati.



- L'interesse principale oggi è rivolto alle classi di linguaggi che includono propriamente $\{CFL\}$ e che non sono inclusi propriamente in $\{CSL\}$.
- E' noto che un CFL può essere sempre riconosciuto in un tempo polinomiale. Invece non è stato trovato ancora un algoritmo polinomiale per il riconoscimento di un qualsiasi CSL . Non è chiaro dove si collochi il confine tra la classe dei linguaggi a riconoscitori P (polinomiali) e la classe dei linguaggi a riconoscitori NP (non polinomiali).

Livello 3

- Il **linguaggio regolare (RL)** è legato al concetto di **insieme regolare (regular set)** o di **espressione regolare (RE)**.
- Def. Tutte e sole le espressioni ottenute applicando un numero finito di volte, in ordine libero, le operazioni di concatenazione, disgiunzione e star di Kleene su elementi di Σ_T sono dette **espressioni regolari su Σ_T**
- Esempio: un'espressione regolare su $\{a,b\}$ è $(a+(a+b)^*aa)(a+b)^*$
- Def. L è **regolare** (o **razionale**) su $\Sigma_T \Leftrightarrow L$ è espresso da una espressione regolare su Σ_T .
- Def. Una grammatica G è **lineare a destra (RLG)** $\Leftrightarrow P = \{A \rightarrow \beta B, \mid A \in \Sigma_N, B \in \Sigma_N \cup \{\varepsilon\}, \beta \in \Sigma_T^*\}$
- Def. G è **lineare a sinistra (LLG)** $\Leftrightarrow P = \{A \rightarrow B\beta, \mid A \in \Sigma_N, B \in \Sigma_N \cup \{\varepsilon\}, \beta \in \Sigma_T^*\}$
- Def. G è una **grammatica regolare (RG)** $\Leftrightarrow G$ è una RLG oppure una LLG
- Le $LLGs$ sono equivalenti alle $RLGs$: data una LLG c'è sempre un modo per costruire una RLG tale che $L(RLG) = L(LLG)$
- Una RLG avrà regole di produzione riconducibili a

$$A \rightarrow a$$

$$B \rightarrow bC$$

- Esempi di linguaggi non regolari:

$$L = \{a^n b^n \mid n > 0\}$$

$$L = \{a^n c b^n \mid n > 0\}$$

$$L = \{a^n b^k \mid n > k \geq 0\}$$

$$L = \{a^n b^k a^k \mid n, k \geq 0\}$$

$$L = \{a^n \mid n \text{ è un numero primo}\}$$

$$L = \{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\}. \quad \text{Oss. } L \supset \{a^n b^n \mid n > 0\}$$

$$L = \{w \in \{a,b\}^* \mid n_a(w) > n_b(w)\}$$

- Ogni FA può solamente riconoscere un linguaggio regolare, e ogni linguaggio regolare è riconoscibile da un FA .

- Il **Teorema di Kleene** dimostra che strutture equivalenti ai *FA* sono le *espressioni regolari* e i **grafi di transizione (TG)**. I *TG* sono una generalizzazione dei *FA*, in quanto per definizione possono avere più di uno stato iniziale e possono essere non deterministici. Il non determinismo si riferisce al fatto che esiste almeno una sequenza di ingresso che porta il *TG* in uno stato finale attraverso cammini alternativi. Il Teorema di Kleene dimostra che per ogni *TG* non deterministico, esiste un *TG* deterministico che accetta lo stesso linguaggio
- Def. Detti Σ_1 e Σ_2 due alfabeti non necessariamente diversi, un **omomorfismo** è una funzione $h: \Sigma_1 \rightarrow \Sigma_2^*$. L'**immagine omomorfica** di un linguaggio $L \subseteq \Sigma_1^*$ è il linguaggio $h(L) \subseteq \Sigma_2^*$ ottenuto sostituendo i simboli a_i di ogni parola $w \in L$ con le stringhe $h(a_i)$.
- Def. Un **morfismo** [**antimorfismo**] è una funzione $\alpha: \Sigma^* \rightarrow \Sigma^* \Leftrightarrow \forall w, v \in \Sigma^* \alpha(wv) = \alpha(w)\alpha(v)$ [$\alpha(wv) = \alpha(v)\alpha(w)$]
- Def. Un **morfismo** [**antimorfismo**] biiettivo α è un **isomorfismo** [**anti-isomorfismo**]
- Def. Detti L_1 e L_2 due linguaggi su Σ , il **quoziente a destra** (*right-quotient*) è il linguaggio $L_1/L_2 = \{w \mid wv \in L_1, v \in L_2\}$. In altre parole dovremo sottrarre da L_1 tutte le stringhe che hanno un suffisso in L_2 .
- Applicando una delle seguenti operazioni sui *RLs* L_1 e L_2 si ottiene un *RL* (si dice che i *RLs* sono chiusi rispetto ad esse)

- $L_1 \cup L_2$ (stringhe appartenenti a L_1 o a L_2)
- $L_1 \cap L_2$ (stringhe appartenenti a L_1 e a L_2)
- $L_1 L_2$ (stringhe con prefisso in L_1 e suffisso in L_2)
- $-L_1$ (tutte le stringhe in $\Sigma^* - L_1$)
- L_1^* (concatenazione di zero o più stringhe di L_1)
- $L_1 - L_2$ (stringhe appartenenti a L_1 ma non a L_2)
- L_1^R (stringhe appartenenti a L_1 riflesse)
- $h(L_1)$ (immagine omomorfica di L_1)
- L_1/L_2 (quoziente a destra)

- La chiusura dei *RLs* sulle operazioni anzidette può essere usata per dimostrare che un dato linguaggio non è regolare. La prova è sempre per assurdo:

- 1) si assume L regolare
- 2) si costruisce un linguaggio che è noto essere non regolare L' applicando le proprietà di chiusura a L e ad altri *RLs*
- 3) per contraddizione, si conclude che L non è regolare

- *Esempio 1. Th.* $L = \{a^n b^n a^n \mid n > 0\}$ non è regolare. *Dim.* Supponiamo L sia un *RL*. Allora, preso il *RL* $L_1 = \{a^*\}$, per chiusura sotto il quoziente a destra, $L_2 = L/L_1 =$

$\{a^n b^n a^k \mid n \geq k \geq 0, n > 0\}$ è regolare. Analogamente, per chiusura sotto intersezione $L_3 = L_2 \cap \{a^* b^*\} = \{a^n b^n \mid n > 0\}$ è regolare. Contraddizione. QED.

- *Esempio 2. Th.* $L = \{a^3 b^n c^{n-3} \mid n > 3\}$ non è regolare. *Dim.* Supponiamo L sia un RL . Allora è regolare l'immagine omomorfica $h(L) \subseteq \{a, b\}^*$ ottenuta sostituendo $h(a)=b, h(b)=b, h(c)=a, L_1 = \{b^{n+3} a^{n-3} \mid n > 3\} = \{b^{n+6} a^n \mid n > 0\}$. Per chiusura sotto riflessione $L_2 = L_1^R = \{a^n b^{n+6} \mid n > 0\}$ è regolare. Preso il RL $L_3 = \{b^6\}$, per chiusura sotto il quoziente a destra, $L_4 = L_2/L_3 = \{a^n b^n \mid n > 0\}$ è regolare. Contraddizione. c.v.d.

- Un RL è dato in **rappresentazione standard** se è specificato in uno dei seguenti modi:
 - a) un FA (DFA o NFA)
 - b) un'espressione regolare
 - c) una RG
- Dato un RL L su Σ_T in rappresentazione standard, c'è sempre un modo per stabilire:
 - a) se una stringa Σ_T^* appartiene a L
 - b) se L è finito, vuoto o infinito
 - c) se L è identico al RL L' , anch'esso dato in rappresentazione standard
- Occorre ricordare che un DFA non è in grado di "contare".
- **Pumping Lemma (PL)** Se L è un RL infinito allora $\exists m > 0 \mid \forall w \in L$ con $|w| \geq m$ si ha $w = xyz$, con $|xy| \leq m, |y| \geq 1$ e $xy^i z \in L$ ($i = 0, 1, 2, \dots$)

In altre parole, se L è regolare può essere espresso da un DFA , con un numero finito $|Q|$ di stati. Dato che L è infinito, alcune stringhe di L avranno lunghezza maggiore di $|Q|$. Per queste stringhe il cammino attraverso il DFA deve contenere un ciclo. Ripetendo il ciclo un numero arbitrario di volte si avrà sempre una stringa accettata dal DFA , cioè appartenente a L .

- Il PL può essere usato per dimostrare che un dato linguaggio non è regolare. La prova è sempre per assurdo:
 - 1) si assume L regolare
 - 2) si mostra che, presa una stringa sufficientemente lunga di L e ripetendo il ciclo un certo numero di volte ("pumping the cycle") si ottiene una stringa che non appartiene a L .
 - 3) si conclude che L non è regolare

Esempio. Th. Supponiamo per assurdo che $L = \{a^n b^n \mid n > 0\}$ sia regolare. *Dim.* Sia m un numero maggiore del numero di stati del DFA che accetta L . Prendiamo $w = a^n b^n$ dove $n > m$. Allora potremo certamente esprimere $w = xyz$, con $|xy| \leq m$. Ora xy sarà composto interamente di a . Secondo il PL L dovrà contenere anche tutte le stringhe $xy^i z$ ($i = 0, 1, \dots$). Ma scegliendo $i = 0$ otteniamo una stringa in cui abbiamo tagliato via tutte le a di y , cioè una stringa che ha meno a di b . Ma questa

stringa non può appartenere a L . Questa contraddizione ci permette di concludere che L non è un RL .

Livello 2

- Una **grammatica indipendente dal contesto** (*context free grammar*, **CFG**), è quella le cui produzioni si applicano indipendentemente dal contesto simbolico in cui si trova l'espressione di partenza,

$$A \rightarrow \gamma \quad (\gamma \in \Sigma^*)$$

- *Def.* L è un **linguaggio indipendente dal contesto (CFL)** $\Leftrightarrow \exists CFG \mid L = L(CFG)$
- Si verifica immediatamente che tutte le regole di produzione di una RG sono regole di produzione valide per una CFG , sicché ogni RL è anche un CFL .
- Quando si fa riferimento alla classe di appartenenza di un linguaggio occorre essere il più specifici possibile. Se dico che L è un CFL , probabilmente voglio dire che L non è un RL , oppure che non so se L sia un RL . Ma se io so che L è un RL allora dovrei chiamarlo un RL , non un CFL .
- Data una CFG , allora sappiamo che $L(CFG)$ è un CFL . Tuttavia ciò non significa che $L(CFG)$ non possa essere un RL . Infatti si potrà dare una RG tale che $L(RG) = L(CFG)$.
- *Esempio.* Data la grammatica

$$G = \langle \{a, b\}, \{S, A, B\}, \{S \rightarrow AB, A \rightarrow aA, A \rightarrow \varepsilon, B \rightarrow Bb, B \rightarrow \varepsilon\}, S \rangle$$

G è una CFG ? Sì. G è una RG ? No. $L(G)$ è un CFL ? Sì. $L(G)$ è un RL ? Sì!

Infatti $L(G)$ può essere generato anche dalla grammatica

$$G = \langle \{a, b\}, \{S, A, B\}, \{S \rightarrow A, A \rightarrow aA, A \rightarrow B, B \rightarrow Bb, B \rightarrow \varepsilon\}, S \rangle$$

che è regolare.

- *Esempio:* $L = \{a^n b^n \mid n \geq 0\}$. La grammatica corrispondente è
 $S \rightarrow aSb$
 $S \rightarrow \varepsilon$
- *Esempio:* $L = \{a^n b^k \mid n > k > 0\}$. La grammatica corrispondente è
 $S \rightarrow aSb$
 $S \rightarrow A$
 $A \rightarrow Aa$
 $A \rightarrow a$
- *Esempio:* $L = \{ww^R \mid w \in \{a, b\}^*\}$. La grammatica corrispondente è
 $S \rightarrow aSa$

$S \rightarrow bSb$

$S \rightarrow \varepsilon$

- *Esempio:* $L = \{w \in \{a,b\}^* \mid n_a(w) = n_b(w)\}$, dove le stringhe di L hanno un eguale numero di a e di b . L non è un RL . Consideriamo la grammatica G

$S \rightarrow aSb$

$S \rightarrow bSa$

$S \rightarrow SS$

$S \rightarrow \varepsilon$

- *Problema:* $L(G) = L$?

- *Esempio:* $L = \{w \in \{(,)\}^* \mid w \text{ è un'espressione bilanciata di parentesi aperte e chiuse}\}$. L è una CSL , ma non è un RL .

$S \rightarrow (S)$

$S \rightarrow SS$

$S \rightarrow \varepsilon$

- A differenza di un FA , un **automa pushdown non deterministico (NPDA)** possiede una memoria di **stack**. Per determinare gli stati interni e di uscita al passo successivo le funzioni di transizione devono tenere conto anche del contenuto corrente dello stack (il top). A ogni passo lo stack può essere aggiornato. Tipicamente questi automi servono ad accettare un linguaggio piuttosto che a simularne la grammatica. Infatti non scrivono sul nastro, e la loro uscita è sostanzialmente il movimento da imprimere al nastro. L'espressione in ingresso è accettata se al termine dell'elaborazione lo stack è vuoto, oppure se l'automa si stabilizza su un determinato stato interno.
- I PDA s ad ogni transizione possono leggere nello stack solo un intero. Questo è sufficiente a consentire loro di confrontare lunghezze di sottosequenze della stringa di ingresso. Per questo i PDA s e i linguaggi da essi riconosciuti sono anche detti "a contatore".
- Formalmente abbiamo che un $NPDA$ è una 7-upla

$$PDA = \langle \Sigma_T, \Gamma, Q, \delta, q_0, z, F \rangle$$

dove Γ è l'**alfabeto di stack** e $z \in \Gamma$ è il **simbolo iniziale dello stack**.

- La funzione di transizione estesa diventa

$$\delta: Q \times \Sigma_T^* \times \Gamma \rightarrow 2^{Q \times \Gamma^*}$$

Come si vede il terzo argomento di δ è il simbolo al top dello stack. L'effetto della transizione è quello di rimuovere il simbolo letto dallo stack. Mentre il simbolo di ingresso può essere il simbolo nullo, dev'essere sempre letto ed eliminato un

simbolo di stack. Se lo stack è vuoto, nessuna transizione è più possibile. Ogni volta che avviene una transizione un nuovo stato viene raggiunto e una nuova stringa di simboli $x \in \Gamma^*$ viene “push-ata” sullo stack. In realtà, dato che abbiamo a che fare con un automa non deterministico, la transizione raggiunge un insieme finito di stati diversi e “push- a ” sullo stack un insieme finito di numeri.

- *Def.* Una grammatica è in **Chomsky Normal Form** se tutte le sue produzioni sono del tipo

$$A \rightarrow BC$$

$$A \rightarrow a$$

- *Th.* Qualsiasi CSG che non contempli la produzione $S \rightarrow \varepsilon$ può sempre espressa nella Chomsky Normal Form

- *Esempio:* Il linguaggio $L = \{a^n b^n \mid n > 0\}$ può essere riscritto

$$S \rightarrow UB$$

$$S \rightarrow AB$$

$$U \rightarrow AS$$

$$A \rightarrow a$$

$$B \rightarrow b$$

- Una grammatica è in **Greibach Normal Form** se tutte le sue produzioni sono del tipo

$$A \rightarrow a\chi \quad (\chi \in \Sigma_N^*)$$

- un **automa pushdown deterministico (DPDA)** differisce da un NPDA nello stesso senso in cui un NFA differisce da un DFA. Tuttavia mentre NFAs e DFAs sono equivalenti, NPDAs e DPDA non sono equivalenti.

- *Def.* L è un **linguaggio deterministico indipendente dal contesto (DCFL)** se è riconosciuto da un DPDA. Si ha che $\{DCFL\} \subset \{CFL\}$.

- *Esempio.* Un DCFL è il **linguaggio di Dyck**, formato da tutte le sequenze bilanciate di parentesi aperte e chiuse. $L = \{w \in \{(,)\}^* \mid w \text{ è un'espressione bilanciata di parentesi aperte e chiuse}\}$

- **Pumping Lemma per CFLs.** Se L è un CFL infinito allora $\exists m > 0 \mid \forall w \in L$ con $|w| \geq m$ si ha $w = uvxy$, con $|uvx| \leq m$, $|ux| \geq 1$ e $u^i v x^i y \in L$ ($i = 0, 1, 2, \dots$)

- Il PL per CFLs può essere usato per dimostrare che un dato linguaggio non è indipendente dal contesto. La prova è sempre per assurdo:

- *Def.* Una variabile A è **ricorsiva** $\Leftrightarrow A$ può generare una stringa contenente A stessa.

Livello 1

- Secondo la definizione classica, una **grammatica dipendente dal contesto (context sensitive grammar, CSG)**, è quella le cui produzioni sono del tipo

$$\alpha A \beta \rightarrow \alpha \gamma \beta \quad (\alpha, \beta, \gamma \in \Sigma^*)$$

Secondo un'altra definizione, una *CSG* è quella con produzioni

$$\alpha \rightarrow \beta \quad (\alpha \in \Sigma^+, \beta \in \Sigma^*, |\alpha| \leq |\beta|)$$

Da questa definizione si vede come una *CSG* è **non-contrante** (*noncontracting*) nel senso che ogni passo di derivazione non può mai accorciare la lunghezza della forma sentenziale

- Def. L è un **linguaggio dipendente dal contesto (CSL)** $\Leftrightarrow \exists \text{ CSG} \mid L = L(\text{CSG})$
- Esempio: Il linguaggio $L = \{abc, aabbcc, \dots\} = \{a^n b^n c^n \mid n > 0\}$ può essere generato dalla grammatica di tipo 1

$$S \rightarrow aSBC$$

$$S \rightarrow aBC$$

$$bB \rightarrow bb$$

$$cC \rightarrow cc$$

$$bC \rightarrow bc$$

$$aB \rightarrow ab$$

$$CB \rightarrow BC$$

- **Teorema di Kuroda**: una qualsiasi grammatica c.s. ...
- Th. $\{CSL\} \supset \{CFL\}$
- Th. $\{RCL\} \supset \{CSL\}$
- Un **automa lineare limitato** (*Linear Bounded Automata, LBA*) è un *FA*, con un apparato di lettura/scrittura che può agire su un nastro di lunghezza $k|w|$ dove w è la sequenza iniziale in ingresso, e k un intero. Le funzioni di transizione hanno come ingresso il simbolo letto sul nastro e lo stato interno, come uscita il simbolo da scrivere sul nastro, lo stato seguente e il movimento da imprimere al nastro stesso. Una sequenza è accettata dall'automa se al termine dell'elaborazione l'automa si stabilizza su un determinato stato interno.
- Th. $\forall CSL \exists LBA \mid CSL = L(LBA)$
- Th. $\forall LBA \exists CSL \mid CSL = L(LBA)$

Livello 0

- Il linguaggi **ricorsivamente enumerabili (REL)** sono quelli di tipo più generale, per i quali esiste una procedura deterministica di generazione. Le **grammatiche non ristrette (UG)** hanno ovviamente la forma più generale che si possa immaginare

$$\alpha \rightarrow \beta \quad (\alpha \in \Sigma^+, \beta \in \Sigma^*)$$

Macchina di Turing

- Alan Turing definì la **macchina di Turing (TM)** nel tentativo di formalizzare la nozione di **procedura effettiva** (*effective procedure*), sostanzialmente lo stesso che dire *algoritmo*.
- Def. Una procedura è *effettiva* (o *meccanica*) \Leftrightarrow
 - a) è fatta di un numero finito di istruzioni simboliche esatte e finite
 - b) produce sempre il risultato voluto in un numero finito di passi
 - c) può in linea di principio essere sempre eseguita da un uomo con solo carta e matita
 - d) non richiede né intuito né inventiva da parte dell'uomo
- La *TM* è un'idealizzazione di un uomo che calcola: "A man provided with paper, pencil, and rubber" and a subject to strict discipline, is in effect a universal machine" (Turing, 1948). "Turing's Machines. These machines are humans who calculate" (Wittgenstein, 1890).
- La *TM* è l'automa più generale e potente finora ideato. Il funzionamento è simile a quello di un *LBA*, solo che il nastro accessibile alle operazioni di lettura/scrittura è infinito, e non limitato a quello occupato dalla stringa di ingresso iniziale (o un suo multiplo). Come per ogni *FA* il numero di stati interni della *TM* deve essere finito ($|Q|=M$).
- Anche la *TM* è provvista di un'apparato di lettura/scrittura che scorre sequenzialmente in una direzione (R) o nell'altra (L) sulle caselle del nastro, leggendo, segnando o cancellando un simbolo appartenente all'alfabeto di ingresso Σ_T . Il nastro è inizialmente vuoto (tutti i simboli *blank*) eccetto, eventualmente, una sequenza finita contigua di caselle, la sequenza di ingresso. In tal caso il dispositivo di lettura è posizionato su uno qualunque dei simboli non vuoti.
- La *TM* utilizza lo stesso nastro per riportare i risultati parziali e il risultato finale, la sequenza di uscita.
- Formalmente una *TM* è una 7-upla

$$T = \langle \Sigma_T, \Gamma, Q, \delta, q_0, \#, F \rangle$$

dove Γ è l'**alfabeto di nastro**, δ la funzione di transizione **parziale**, $\# \in \Gamma$ è il simbolo **blank**.

- Molti libri di testo non fanno distinzione tra Σ_T e Γ . Il fatto è che sul nastro potrebbero esserci simboli diversi e più numerosi rispetto a quelli della sequenza di ingresso (p.es. $\# \in \Gamma$, ma $\# \notin \Sigma_T$). Poniamo $\Gamma \supset \Sigma_T$.

- La funzione di transizione è

$$\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{L, R\}$$

Quindi, quando la *TM* è in uno stato (q) e legge un certo simbolo (γ) sul nastro, lo sovrascrive con qualche altro simbolo (γ'), va in un qualche altro stato (q') e muove la testina di lettura/scrittura sul nastro una posizione a sinistra (L) o a destra (R).

- Una **descrizione istantanea** di una *TM* deve dirci (1) lo stato in cui si trova, (2) il contenuto del nastro, (3) la posizione della testina sul nastro. Un esempio è il seguente

$$x_i \dots x_j q_m x_k \dots x_l$$

dove le x sono i simboli presenti sul nastro a sinistra della testina, q_m è lo stato corrente, x_k è il simbolo sotto la testina, le rimanenti x sono i simboli presenti sul nastro a destra della testina

- La mossa di una *TM* può essere quindi rappresentata da una coppia di descrizioni istantanee, separate dal simbolo \mid . P.es., se una transizione ammessa è

$$\delta(q_5, b) = (q_8, c, R)$$

una mossa potrebbe essere

$$abbabq_5babbb \mid abbaabcq_8abbb$$

Programmare una *TM* significa dare una sequenza di 5-uple

$$(stato-corrente, simbolo-letto, prossimo-stato, simbolo-scritto, direzione)$$

o anche

$$stato-corrente \quad simbolo-letto \rightarrow prossimo-stato \quad simbolo-scritto \quad direzione$$

- Spesso una *TM* è definita avere la testina inizialmente posizionata sul primo simbolo non vuoto all'estrema sinistra della stringa di ingresso. In realtà ciò non è necessario, infatti la *TM*, partendo da una qualsiasi casella non vuota, può portarsi sulla prima eseguendo il semplice programma

$$\begin{aligned} &(q_0, a, q_0, a, L) \\ &(q_0, b, q_0, b, L) \\ &(q_0, \#, q_1, \#, R) \end{aligned}$$

dove si è assunto $\Sigma_T = \{a, b\}$.

Una TM si arresta quando non ha più mosse possibili. Se si arresta su uno stato finale allora accetta la stringa di ingresso, altrimenti no. Quindi

- Def. $L(TM) = \{w \in \Sigma_T^+ : q_0 w \vdash^* u q_f v \text{ con } q_f \in F, u, v \in \Gamma^*\}$
- (la precedente definizione assume che la TM parta con la testina posizionata sul simbolo più a sinistra di w)
- Una TM può non arrestarsi per due motivi:
 - a) termina in uno stato non finale
 - b) termina in un loop infinito
- Def. Una **computazione** è la sequenza di configurazioni istantanee di una TM dal momento di inizio al momento di arresto in uno stato finale
- Esempio. Il linguaggio $L = \{a^n b^n \mid n > 0\}$ può essere riconosciuto dalla seguente TM, avente stati $Q = \{q_0, \dots, q_4\}$ di cui q_0 è iniziale e q_1 finale.

$(q_0, a, q_0, a, L); (q_0, b, q_0, b, L); (q_0, \#, q_1, \#, R)$	portati tutta a sinistra
$(q_1, a, q_2, \#, R)$	cancella una a
(q_1, b, q_4, b, R)	se c'è una b vai in q_4 e stop
$(q_2, a, q_2, a, R); (q_2, b, q_2, b, R); (q_2, \#, q_3, \#, L)$	portati tutta a sinistra
$(q_3, b, q_0, \#, L)$	cancella una b
$(q_3, \#, q_4, \#, L)$	se c'è una $\#$ vai in q_4 e stop

- Una TM può essere usata come trasduttore (*transducer*) di sequenze in ingresso in sequenze di uscita.
- Def. Una funzione $y = f(x)$, $x, y \in \Sigma^*$, è **T-computabile** $\Leftrightarrow \exists TM : q_0 x \vdash^* q_f y$ con $q_f \in F$

Variazioni sulle TMs

- Th. La classe delle TMs che hanno la possibilità di non spostare la testina sul nastro al termine di una mossa è equivalente alla classe delle TMs standard
- Th. La classe delle TMs che leggono un intero vettore di simboli alla volta su ciascuna casella (**N-trak TMs**) è equivalente alla classe delle TMs standard
- Th. La classe delle TMs che hanno un nastro semi-infinito è equivalente alla classe delle TMs standard
- Th. La classe delle TMs che hanno due nastri, uno di sola lettura che contenente l'input iniziale e l'altro di lettura/scrittura inizialmente vuoto (**Off-line TMs**), è equivalente alla classe delle TMs standard
- Th. La classe delle TMs che hanno un numero finito di nastri indipendenti (**multiTape TMs**) è equivalente alla classe delle TMs standard
- Th. La classe delle TMs in cui il nastro è controllato da un NFA (**Nondeterministic TMs**) è equivalente alla classe delle TMs standard
- Th. La classe delle TMs che leggono/scrivono e si muovono su spazi discreti multidimensionali (**N-dimensional TMs**) è equivalente alla classe delle TMs standard

- Th. La classe delle *TMs* che hanno solo 2 simboli di nastro (**binary TMs**) è equivalente alla classe delle *TMs* standard
- Th. La classe delle *TMs* che hanno solo 2 stati (**2 state TMs**) è equivalente alla classe delle *TMs* standard

TM universale

- Def. Un dispositivo computazionale è universale \Leftrightarrow può simulare un qualsiasi altro dispositivo computazionale.
- Un modo di dimostrare che un dispositivo computazionale è universale è quello di fargli simulare un altro dispositivo computazionale universale noto
- Un automa è necessariamente cablato per eseguire un certo algoritmo, sempre lo stesso. Una **TM universale (UTM)** è un automa capace di emulare l'azione di una qualsiasi *TM* su una sequenza di ingresso data. Una volta che sia stata opportunamente codificata, la *TM* che dev'essere emulata è a tutti gli effetti un **programma**, da dare in ingresso alla *UTM*. L'*UTM* può essere interpretata come un computer *general purpose*.

Codifica di una TM

- Memorizziamo il programma come una sequenza di 5-uple:

stato-corrente simbolo-letto \rightarrow prossimo-stato simbolo-scritto direzione

- La *TM* che dev'essere emulata può avere un alfabeto più numeroso dell'alfabeto della *UTM*. Una codifica conveniente per i simboli dell'alfabeto da emulare, $\Sigma_T = \{a_j\}$, è la codifica unaria, $a_1=1, a_2=11, a_3=111, \dots$. Il simbolo "0" può essere usato come separatore. Un'altra possibile codifica è la codifica binaria, $a_1=1, a_2=10, a_3=11, \dots$, ma dovremo introdurre un terzo simbolo "x" come separatore.
- Codificando allo stesso modo l'indice degli stati q_j e simboli speciali "R", "L", arriviamo a rappresentare la *TM* con una sequenza finita $[TM]$, del tipo $[TM] = 101110..01$, cioè in definitiva un numero naturale n (p.es. secondo la notazione binaria). Inversamente, per ogni numero $n = 0, 1, \dots$ è definita una *TM*, T_n . Anche la stringa dei dati in ingresso e quella dei dati in uscita corrisponderanno ai numeri interi m e p , sicché $T_n(m) = p$.
- Un esempio di codifica è

$$q_i a_h \rightarrow q_j a_k m \quad \Leftrightarrow \quad 1^i 0 1^h 0 1^j 0 1^k 0 1^m$$

con $i, j = 1, \dots, M$; $m = 1, 2$ ("L" $\Leftrightarrow m=1$, "R" $\Leftrightarrow m=2$). Le varie quintuple possono essere concatenate una dopo l'altra, senza un ordine particolare, separandole con un 0^2 .

Esempio:

TM:

$$\begin{array}{lll}
 q_1 0 \rightarrow q_2 1 L & \Leftrightarrow & 10101101101 \\
 q_1 1 \rightarrow q_3 0 L & \Leftrightarrow & 101101110101 \\
 q_2 0 \rightarrow q_3 1 L & \Leftrightarrow & 1101011101101 \\
 q_2 1 \rightarrow q_2 0 L & \Leftrightarrow & 110110110101
 \end{array}$$

Questa *TM* riconosce le parole del linguaggio $L = \{1, 01^*0\}$ presenti su un lato del nastro e le riscrive invertite sul lato opposto. Può essere rappresentata dal numero binario

$$[TM] = 101011011010010110111010100110101110110100110110110101000$$

oppure dal numero binario

$$[TM] = 101101110101001010110110100110110110101001101011101101000$$

Ogni sequenza binaria di tipo (TM)

$$\{1^i 0 1^h 0 1^j 0 1^k 0 1^m 00\} * 0$$

codifica una sola *TM*, mentre una data *TM* può essere codificata da molte sequenze binarie di tipo (TM). Per definizione, le sequenze binarie che non hanno struttura (TM) codificano *TM*s che riconoscono il linguaggio vuoto $\{\varepsilon\}$, cioè che non accettano nessuna parola in $\{0, 1\}^*$.

- Vi è quindi la possibilità di costruire una *UTM*, U , tale che per ogni n e m , $U(n, m) = T_n(m)$. Anche U sarà associata a un numero u , cioè $U = T_u$.
- Per quanto riguarda la realizzabilità di una *UTM* pensiamo a una *TM* con 3 nastri. Il primo nastro contiene il programma, sotto forma di una sequenza di 5-uple del tipo già descritto, un secondo nastro memorizza lo stato corrente della *TM* da emulare T , e il terzo nastro contiene la sequenza di ingresso m (all'inizio della computazione) e di uscita p (al termine della computazione). A questo punto noi sappiamo che una *TM* con 3 nastri è equivalente a una *TM* con un solo nastro. Tutto sommato una *UTM* non è così complicata come si potrebbe pensare.

RAM

- Una **Macchina ad Accesso Casuale** (*Random-Access Machine*, **RAM**) è definita dai seguenti quattro oggetti
 - a) **Tipi di dati.** solo interi senza segno (0, 1, 2, 3, ...)
 - b) **Variabili.** Un numero arbitrario di variabili inizializzate a 0
 - c) **Test.** Può solo verificare la condizione $\langle \text{variabile} \rangle = 0$
 - d) **Istruzioni:**

- **If** <test> **then** <istruzione> **else** <istruzione>;
- **While** <test> **do** <istruzione>;
- <variabile> := <variabile> + 1
- <variabile> := <variabile> - 1 (il decremento di una variabile a 0 non ha effetto)

Inoltre è permessa l'esecuzione di più istruzioni in sequenza separate da “;”, e l'uso di parentesi per raggruppare una sequenza di istruzioni in un'unica istruzione.

Si dimostra che questo semplice linguaggio è equivalente in potenza alla TM.

La tesi di Church-Turing

- **Tesi di Turing:** qualsiasi computazione effettiva può essere eseguita da una *TM*.
- Oss. L'implicazione inversa è ovvia
- Church arrivò allo stesso risultato qualche mese prima di Turing, usando il concetto di “funzione lambda-definibile” invece di “funzione T-computabile”. Però Church restrinse il campo al calcolo delle sole funzioni di numeri interi.
- **Tesi di Church:** una funzione di interi positivi è effettivamente calcolabile \Leftrightarrow è ricorsiva
- Fu Turing stesso a dimostrare l'equivalenza tra le due classi di funzioni nel 1936.
- Identico è pure il concetto di “funzione ricorsiva” (Godel, 1934; Herbrand, 1932), come dimostrarono Church e Kleene (1936). Queste dimostrazioni si riferivano sempre al caso di funzioni di numeri interi. Per queste funzioni Kleene riconobbe nel 1967 l'equivalenza tra le tesi di Church e di Turing, proponendo la denominazione *Tesi di Church-Turing*.
- E' dimostrato che i seguenti modelli formali di calcolo sono equivalenti l'uno all'altro:
 - Alan M. Turing – *Macchine di Turing*
 - Alonzo Church – *Lambda calculus*
 - Emil Post (1943, 1946) – *sistemi di produzione normali e canonici*
 - Raymond Smullyan – *sistemi formali*
 - Stephen Kleene – *teoria delle funzioni ricorsive*
 - Noam Chomsky – *grammatiche non ristrette*
 - Schonfinkel (1924) – *definibilità combinatoria*
 - Markov (1969) – *algoritmi di Markov*
- Molte interpretazioni erronee della tesi di Turing si sono avute in letteratura. Ad esempio la ***Tesi di Turing in forma debole***: una *TM* può calcolare qualunque funzione che possa essere calcolata da un computer con architettura qualsiasi

(Dennett, 1991). Oppure la **Tesi di Turing in forma forte**: tutte le funzioni calcolabili sono T-computabili.

- La **Tesi di Turing in forma forte** è indimostrabile in quanto fa appello a una nozione “metamatematica” di computabilità. Tuttavia, in linea di principio la tesi sarebbe confutabile, se si trovasse un qualche algoritmo non implementabile su una TM. **Pour-El e Richards** (1983) hanno trovato un algoritmo simile, ma la cosa è controversa.

Il problema dell'arresto

- **Alan M. Turing** nacque a Londra nel 1912. Studiò a Cambridge. Nel 1936 pubblicò *On Computable Numbers, with an Application to the Entscheidungsproblem*, dove definiva la teoria delle *macchine di Turing* e dimostrava l'indcidibilità del *problema dell'arresto* (*Halt Problem*). Il “problema della decisione” (*Entscheidungsproblem*) era stato proposto da Hilbert al Congresso Internazionale dei Matematici a Parigi (1928), come **decimo problema di Hilbert**:

Assegnata una certa classe di problemi matematici, esiste un procedimento meccanico generale capace, in linea di principio, di risolverli tutti?

oppure (versione di Church):

Esiste un metodo effettivo che, data un'espressione P nella notazione di un sistema di logica simbolica (calcolo dei predicati) stabilisca se P è dimostrabile o no?

Turing dimostrò che non esiste una TM che possa decidere, in un numero finito di passi, se una data proposizione è vera o falsa. Quindi, in base alla tesi di Turing, non esiste metodo effettivo per fare la stessa cosa. Turing era convinto difensore dell'AI. Inventò con l'amico D. Champenowne il primo programma per giocare a scacchi (*Turochamp*). Morì a 41 anni per un'intossicazione da farmaci o per suicidio.

- **problema dell'arresto (Halting Problem)** Il X problema di Hilbert poteva essere riformulato nel problema se esista una TM capace di decidere se una qualsiasi TM si arresti dopo un tempo finito ($TM \downarrow$) o no ($TM \uparrow$). La risposta di Turing fu no. La dimostrazione consiste in una *reductio ad absurdum* partendo dall'ipotesi che una tale TM esista, chiamiamola *H*, e sfrutta la **tecnica diagonale di Cantor**. Sia $H(n,m) \downarrow$ tale che

$$H(n,m) = 1 \text{ se } T_n(m) \downarrow, 0 \text{ altrimenti} \quad (*)$$

Definiamo ora l'algoritmo, che sicuramente si arresta,

$$Q(n,m) = H(n,m)T_n(m) \text{ se } n \neq m, H(n,n)T_n(n) + 1 \text{ altrimenti}$$

Ora, $Q(n,n)$ è una TM , diciamo $T_q(n)$. Ma allora

$$T_q(q) = Q(q,q) = H(q,q)T_q(q) + 1 = T_q(q) + 1$$

e la contraddizione è palese. *QED*.

Un altro modo di presentare è il seguente:

$$H(n,m) \downarrow \text{ se } T_n(m) \uparrow \text{ (condizione meno restrittiva rispetto alla (*))}$$

Posto $m=n$

$$H(n,n) \downarrow \text{ se } T_n(n) \uparrow$$

Ora, $H(n,n)$ è una TM , diciamo $T_q(n)$. Ma allora

$$T_q(q) = H(q,q)$$

con

$$T_q(q) \downarrow \text{ se } T_q(q) \uparrow$$

Dato che ciò è impossibile, ne deduciamo che in realtà $T_q(q) \uparrow$. Quindi esiste un algoritmo che sicuramente non si arresta, e sul quale neanche H si arresta. Quindi H non è in grado di stabilire che $T_q(q) \uparrow$, anche se noi sappiamo per certo che così deve essere.

Decidibilità

- Def. L è **ricorsivamente enumerabile (REL)** $\Leftrightarrow \exists T \mid \forall w \in \Sigma^* (w \in L \Leftrightarrow T(w) \downarrow \text{ e } T(w)=1)$
- Cioè per un REL esiste sempre una TM che si arresta e accetta ogni sua stringa, ma può arrestarsi o non arrestarsi per ogni stringa non sua
- Def. L è **ricorsivo (RCL)** $\Leftrightarrow \exists T \mid \forall w \in \Sigma^* (w \in L \Leftrightarrow T(w) \downarrow \text{ e } T(w)=1) \text{ e } (w \notin L \Leftrightarrow T(w) \downarrow \text{ e } T(w)=0)$
- Cioè per un RCL esiste sempre una TM che si arresta e accetta ogni sua stringa, e si arresta e non accetta ogni stringa non sua
- Nota: la parola “ricorsivo” non deve essere intesa in senso letterale, è solo una terminologia presa a prestito dalla *teoria delle funzioni ricorsive* di Kleene (una TM è equivalente a una *funzione ricorsiva generale*).
- Th. $\{REL\} \supset \{RCL\}$
- Per i REL che non sono RCL non sappiamo in anticipo se la TM si fermerà o no in corrispondenza di stringhe non appartenenti ad essi. Ciò significa che *il problema dell’arresto di una TM è indecidibile (Halting Problem)*

- Def. un **problema sì/no**, **ynP**, è una funzione deterministica $f: A \rightarrow \{0,1\}$, con A insieme di oggetti o relazioni. P.es. il **Primality Problem (problema dei numeri primi)** è uno di questi, con $A=N$, $f(n)=1$ se n è primo, $f(n)=0$ se n non è primo.
- Def. **ynP** è **decidibile** (o **risolvibile**) $\Leftrightarrow \exists T \mid \forall a \in A: T(a) \downarrow \text{ e } T(a)=f(a)$
- Def. **ynP** è **semidecidibile** (o **semirisolvibile**) $\Leftrightarrow \exists T \mid \forall a \in A \mid f(a)=1: T(a) \downarrow \text{ e } T(a)=1$
- Dato uno **ynP**, possiamo definire un alfabeto Σ^* e un linguaggio $L_P \subseteq \Sigma^*$, formato da stringhe che codificano tutti e soli gli elementi $a \in A \mid f(a)=1$. P.es. per il problema dei primi si può scegliere $\Sigma=\{0,1\}$ e $L_P=\{01^n \mid n \text{ è primo}\}$.
- Dalle precedenti definizioni segue che
 - 1) **ynP** è **decidibile** $\Leftrightarrow \exists \Sigma, L_P \subseteq \Sigma^* \mid L_P$ è **ricorsivo**
 - 2) **ynP** è **semidecidibile** $\Leftrightarrow \exists \Sigma, L_P \subseteq \Sigma^* \mid L_P$ è **r.e.**
- Th. Il complemento Σ^*-L di un linguaggio ricorsivo L è ricorsivo. Dim. Se T è la TM che accetta L , la TM che riconosce Σ^*-L è la stessa, con le uscite negate.
- L'unione di due linguaggi ricorsivi (RCL) è un linguaggio ricorsivo (RCL)
- **Teorema di Post.** L e Σ^*-L sono **REL** \Rightarrow è **RCL**. Dim. dette T_1 e T_2 le macchine di Turing che riconoscono L e Σ^*-L rispettivamente, $\forall w \in \Sigma^*$ è sufficiente far eseguire un passo alla volta alternativamente a T_1 e T_2 . La prima che si arresta fornisce la risposta corretta
- Σ^*-L_P non è **REL** $\Rightarrow L_P$ non è **RCL** \Leftrightarrow **ynP** è **indecidibile**.
- Dato $L \subseteq \{0,1\}^*$, riconosciuto da una TM T , $\exists T_1$ con alfabeto di nastro $\{0,1,\#\}$ che riconosce L .
- Collegati al problema dell'arresto sono i seguenti altri **ynP** non decidibili:
 - 1) Stabilire se una qualsiasi TM si arresta quando riceve in ingresso una sequenza infinita di caselle vuote. $A=\{T_j\}$, $f(T)=1$ se $T(..\#..) \downarrow$, $f(T)=0$ se $T(..\#..) \uparrow$.
 - 2) Stabilire se due qualsiasi TM accettano lo stesso linguaggio. $A=\{T_i, T_j\}^2$, $f(T_i, T_j)=1$ se $L(T_i)=L(T_j)$, $f(T_i, T_j)=0$ altrimenti.
 - 3) Stabilire se per qualsiasi TM e K intero positivo, $L(TM)$ contiene sequenze di lunghezza K .
 - 4) Stabilire se per qualsiasi TM, il $L(TM)$ contiene due sequenze di uguale lunghezza.
 - 5) Stabilire se una qualsiasi TM passerà per un determinato stato
 - 6) Stabilire se una qualsiasi TM se $L(TM)$ è infinito
 - 7) Stabilire se una qualsiasi TM se $L(TM)$ è vuoto
- Data una grammatica G context-free è **decidibile se**:
 - a) $L(G)$ è vuoto, finito o infinito
 mentre è **indecidibile se**:
 - b) $L(G)$ è infinito regolare
 - c) G è ambigua
 - d) $\Sigma_T^* - L(G)$ è finito, infinito regolare o context-free
 - e) $L(G_1) = L(G_2)$ (G_1 e G_2 context-free)
 - f) $L(G_1) \subseteq L(G_2)$

- g) $L(G_1) \cap L(G_2)$ è vuoto, finito, infinito regolare o context-free
- **Problema di corrispondenza di E.L. Post.** Date due sequenze di stringhe su Σ (w_1, w_2, \dots, w_m) e (v_1, v_2, \dots, v_m) , stabilire se esiste una sequenza di interi n_1, n_2, \dots, n_k ($k \geq 1$) compresi tra 1 e m tale che $w_{n_1} w_{n_2} \dots w_{n_k} = v_{n_1} v_{n_2} \dots v_{n_k}$.
 - **Esempio.** Siano date le due sequenze $w = (a, abaaa, ab)$ e $v = (aaa, ab, b)$. Allora una sequenza di interi come quella richiesta sarebbe 2, 1, 1, 3
 - Dal teorema di indecidibilità del problema dell'arresto segue il teorema di indecidibilità del problema di corrispondenza di Post.

Insiemi enumerabili

- **Def.** Un insieme è **enumerabile** \Leftrightarrow è possibile porre ogni suo elemento in corrispondenza biunivoca con l'insieme dei numeri naturali
- Il numero totale dei numeri naturali $N = \{0, 1, 2, \dots\}$, il numero totale dei numeri interi $Z = \{\dots, -2, -1, 0, 1, 2, \dots\}$, il numero totale dei numeri razionali $Q = \{p/q, p \text{ e } q \text{ interi primi tra loro, } q \neq 0\}$, sono lo stesso numero, **alef zero** \aleph_0 (*aleph-null*)
- $\mathbb{R} = Q \cup I$; l'insieme dei numeri reali è uguale all'unione dei numeri razionali e irrazionali
- La cardinalità dei numeri reali è $|\mathbb{R}| = 2^{|\mathbb{N}|} = \aleph_1$.
- In generale $\aleph_{i+1} = 2^{\aleph_i}$. Che tra \aleph_i e \aleph_{i+1} non esistano altri ordini di infinito è cosa ancora da dimostrare (**ipotesi del continuo**).
- Tutti i numeri reali computabili sono enumerabili (p.es. basta numerare le *TM* che li generano)
- **Th.** $\sqrt{2}$ è irrazionale
- **Dim.** se per assurdo fosse $\sqrt{2} = p/q$ avremmo $2 = p^2/q^2 \Rightarrow p^2 \text{ è pari} \Rightarrow p^2 = 2p' \Rightarrow 2q^2 = 4p'^2 \Rightarrow q^2 = 2p'^2 \Rightarrow q \text{ è pari} \Rightarrow p \text{ e } q \text{ non sono primi tra loro. (c.v.d.)}$
- L'insieme dei numeri trascendentali (p.es. e, π) è non enumerabile
- L'insieme delle funzioni da N a N è non enumerabile
- L'insieme delle funzioni da A a B è $|B|^{|A|}$.
- L'insieme delle funzioni da N a $\{0, 1\}$ è non enumerabile, e pari a $2^{|\mathbb{N}|}$ cioè $|\mathbb{R}|$
- L'insieme delle $\{TM\}$ è enumerabile
- Dato che i *RELS* sono quelli le cui stringhe sono accettate da una *TM*, l'insieme $\{REL\}$ è numerabile, cioè $|\{REL\}| = \aleph_0$.
- Il powerset di $\{REL\}$, l'insieme di tutti i linguaggi che sono sottoinsiemi di $\{REL\}$, non è enumerabile, infatti ha cardinalità $\aleph_1 = 2^{\aleph_0}$. Quindi esistono infiniti linguaggi che non sono computabili da una *TM*. Chiamiamo questi **linguaggi non ricorsivamente enumerabili**.
- Secondo la tesi di Turing, una *TM* può computare qualsiasi procedura effettiva. Quindi esistono infiniti linguaggi che non possono essere definiti da una procedura effettiva.
- Possiamo trovare un linguaggio non ricorsivamente enumerabile per diagonalizzazione.

INFLUENZE SULLA AI E SULLA CS

- La Teoria Astratta degli Automi ha influenzato la CS e l'AI, nelle sue conseguenze logiche e nei suoi fraintendimenti
- Noam Chomsky si proponeva come anti-empirista, contro i behavioristi e a sostegno della teoria innatista della competenza linguistica. Coerentemente con questa impostazione tese a enfatizzare il ruolo delle grammatiche di basso livello, quelle regolari e quelle indipendenti dal contesto (cioè dall'ambiente). Postula infatti l'esistenza di una **struttura linguistica profonda**, con regole di produzione (o **regole di riscrittura**) del tipo

$$F \rightarrow SN \cdot SV$$

- dove la frase F produce la concatenazione di due simboli non terminali, un **sintagma nominale** SN (soggetto e suoi attributi) e un **sintagma verbale** SV (azione e sue caratterizzazioni). La regola ipotizzata da Chomsky è proprio in **Chomsky Normal Form**.
- La teoria degli automi di Turing ha indotto a credere che comportamenti intelligenti di complessità qualunque potessero essere prodotti da computer aventi una memoria sufficientemente ampia e poche regole di calcolo.

TEORIA DELLA COMPLESSITA'

- Tra le teorie della Complessità, la Teoria dei Sistemi Dinamici, La teoria del Caos, la Teoria dei Processi Autoorganizzanti, la Sinergetica, la Teoria Quantistica della Computazione, Artificial Life.
- La CS studia oggetti estremamente complessi
- Negli anni '60 Edward Lorenz scoprì i sistemi caotici deterministici.
- In Biologia il problema della complessità fu preso in considerazione fin dall'800.
- Psicologi come James J. Gibson e Jean Piaget si ispirarono a concetti biologici
- Negli anni '80 si scoprirono processi caotici nei tracciati EEG e nei EP.
- Una branca della CS è l'Artificial Life, che tra l'altro studia come funzioni cognitive possano emergere in popolazioni di elementi artificiali che devono adattarsi al loro ambiente.
- Storicamente, la Psicologia e la Teoria della Computazione esordirono sotto la concezione meccanicistica, che tendeva a ridurre al semplice il complesso. Ad es. per i comportamentisti la complessità derivava solamente dall'elevato numero di associazioni S-R, non dalla loro struttura intrinseca.
- Nell'ambito della teoria della computazione, la **Teoria della complessità (CT)** si riferisce a due tipi di misura:

La **complessità spaziale (SC)** è una misura della quantità di memoria richiesta. Per una *TM* questo potrebbe coincidere con la lunghezza del nastro usato per eseguire un calcolo. Per un computer digitale potrebbe essere la memoria usata, in *bytes*.

Entrambe queste misure sono una funzione di un singolo parametro di ingresso, la **dimensione dell'input, n** , che può essere misurata in caselle o *bytes*.

La *SC* introdotta da Kolmogorov, Chaitin e Solomonoff alla metà degli anni '60, considera sistemi S che ammettono una rappresentazione in forma di sequenza numerica ω . ω potrà essere l'output di programmi $prog_j$ di lunghezza L_j (p.es. in *bit*).

Def. $SC = \min L_j = L_{min}$

Si dimostra che la scelta del linguaggio di programmazione non influenza *SC*, a meno di una costante additiva, per tutte le stringhe ω a meno di un numero finito.

Diremo che la complessità di S è elevata se $L_{min} \approx |\omega|$ (in tal caso ω è detta **incomprimibile**), bassa se $L_{min} \ll |\omega|$.

Esiste una seconda definizione di complessità spaziale, più generale, che considera oltre alla lunghezza del programma anche la quantità di memoria usata durante l'esecuzione.

La **complessità temporale (TC)** è una misura del tempo richiesto per eseguire un calcolo. Per una *TM* questo potrebbe coincidere col numero di mosse richiesto per eseguire un calcolo. Per un computer digitale potrebbe essere il numero di cicli macchina.

Per una data dimensione, inputs differenti possono avere complessità diverse. Per questo motivo si è soliti determinare la complessità in media e, più spesso, nel **caso peggiore (worst case)**.

La *TC* (o **profondità logica**) è il numero di passi elementari di calcolo perché $prog_{min}$ generi ω . Una sequenza ω casuale ha massima complessità spaziale, ma bassissima complessità temporale. Invece la sequenza delle prime n cifre di π si può ottenere con programmi estremamente semplici, ma ad elevata profondità logica.

- I comportamentisti ipotizzarono che la complessità del comportamento fosse sostanzialmente di tipo spaziale. Anche la Teoria Astratta della Computazione di Turing è sostanzialmente logica, e tende a trascurare gli aspetti temporali della complessità degli algoritmi. La *TM* ha una memoria illimitata e un processore centrale estremamente semplice. Tale impostazione ha diffuso la sensazione che la simulazione di comportamenti intelligenti fosse possibile, pur di avere una memoria sufficientemente ampia.
- Se un algoritmo richiede esattamente $TC(n) = 5n^3 + 2n^2 - n + 1003$ cicli macchina, si semplifica in $TC(n) = O(n^3)$ (ordine n -cubo), cioè trascuriamo i termini di ordine inferiore al massimo e il coefficiente moltiplicativo di quest'ultimo. Il fatto è che i problemi computazionali sono significativi per n grande.
- Def. Un algoritmo è **tempo-polinomiale** $\Leftrightarrow \exists c_0, \dots, c_k \mid TC(n) \leq \sum_{i=0, \dots, k} c_i n^i$

- **Esempio.** Trovare l'elemento massimo di un vettore di lunghezza n richiede $O(n)$ passi
- **Esempio.** Il **Bubble Sort** è un algoritmo di ordinamento che richiede $O(n)$ passi nel caso migliore, ma $O(n^2)$ nei casi medio e peggiore.
- **Esempio.** L'**Heapsort** richiede in media $O(n \log n)$ passi, $O(n^2)$ nel caso peggiore.
- **Esempio.** L'inversione di una matrice $n \times n$ richiede in media $O(n^3)$ passi. Ma oggi si conoscono algoritmi che risolvono il problema in $O(n^{8/3})$ passi.
- **Def.** Un algoritmo è **tempo-logaritmico** $\Leftrightarrow \exists a \mid TC(n) \leq a + \log n$
- Gli algoritmi tempo-logaritmici e tempo-polinomiali sono anche detti problemi **trattabili**, o appartenenti alla classe di complessità **P**.
- La base del logaritmo è irrilevante, dato che la differenza è un fattore costante. Inoltre anche se $n \log n$ non è un polinomio né un logaritmo, il problema è ugualmente trattabile, essendo $n \log n < n^2$.
- I problemi **intrattabili** hanno TC esponenziale, come $O(k^n)$, $O(n!)$, $O(n^n)$
- **Def.** Un algoritmo è **tempo-polinomiale non deterministico**, o **NP** \Leftrightarrow può essere eseguito in un tempo polinomiale su una TM non deterministica, ma sembra essere intrattabile su una TM deterministica
- **Def.** Una TM non deterministica è una macchina capace di indovinare la soluzione di un problema se tale soluzione esiste. Tutto ciò che resta da fare è verificare che la soluzione sia effettivamente tale.
- **Th.** Due problemi NP, X e Y , sono **NP-completi** \Leftrightarrow sono riducibili l'uno all'altro, cioè:
 1. esiste un algoritmo tempo-polinomiale per esprimere X in termini di Y
 2. esiste un algoritmo tempo-polinomiale per trasformare una soluzione di Y in una soluzione di X
- I seguenti problemi sono **NP-completi**:
 1. **Il problema del commesso viaggiatore** (*Travelling Salesman Problem*, **TSP**).
Un commesso viaggiatore parte da Harrisburg e vuole visitare tutte le capitali dei 48 stati uniti d'America, per poi ritornare ad Harrisburg. Stabilire in quale ordine egli dovrà visitare le capitali al fine di minimizzare la distanza totale percorsa. Formalmente, dato un insieme di punti S , una funzione distanza $d: S^2 \rightarrow \mathbb{N}$, un punto iniziale $A \in S$ e un intero k . Determinare tutti i cammini che a partire da A tornano in A dopo aver toccato tutti i punti di S , con una lunghezza totale $\leq k$.
 2. **Il problema del circuito Hamiltoniano** (*Direct Hamiltonian Problem*, **DHP**).
Ogni capitale ha almeno un volo diretto verso un'altra capitale. Trovare, se esiste, la sequenza dei voli che permetta al commesso viaggiatore di visitare tutte le capitali una sola volta.
 3. **Il problema di colorazione** 3. Dato un grafo G indiretto (senza archi bidirezionali) e tre colori a disposizione, decidere se ciascun vertice di G può essere colorato in modo tale che non esistano due vertici dello stesso colore uniti da un arco.

4. **Il problema SAT (Satisfiability Problem).** Data un'espressione booleana ϕ non necessariamente in *CNF* (Conjunctive Normal Form) in n variabili, m connettivi booleani (*and* o *or*) e un numero qualsiasi di *not*, decidere se c'è un'assegnazione delle variabili per la quale ϕ è vera.
 5. **Equivalenza tra espressioni regolari.** Date due espressioni regolari, stabilire se rappresentano lo stesso linguaggio.
 6. **Intersezione di FAs.** Dato un insieme di FAs su Σ , M_1, \dots, M_N , stabilire se esiste una sequenza in Σ^* che è riconosciuta da tutti.
- Se si trovasse un algoritmo tempo-polinomiale per uno solo dei problemi *NP-completi*, lo si troverebbe per tutti gli altri. Finora nessuno c'è riuscito, sicché è probabile (ma non dimostrato) che sia $P \neq NP$.

IL PROBLEMA DELLA MENTE

- La filosofia ha dato storicamente due soluzioni al problema della mente nella natura, quella **dualista** e quella **materialista**.
- La soluzione dualista fu adottata da Platone e da Cartesio. Sin dal '700 l'identificazione del concetto di causalità con quello fisico proprio della meccanica newtoniana ha condotto in un vicolo cieco il problema mente-corpo, aprendo la strada al materialismo. Tuttavia asserire che anche gli stati mentali sono stati fisici non ci dice nulla sulla loro natura intenzionale. Attualmente si cerca di capire quali proprietà degli stati fisici noti possano permetterci di attribuire loro qualità intenzionali e coscienti. Possiamo collocare le teorie della mente in uno spazio a due dimensioni, il peso dell'esperienza e il livello di coscienza.

		Ruolo dell'esperienza	
		Teorie razionaliste	Teorie empiriste
Ruolo della coscienza	Tutti gli stati mentali sono coscienti		<i>Teoria di Searle</i>
	Esistono stati mentali non coscienti	<i>Teoria di Fodor, Teoria di Chomsky</i>	<i>Teoria di Newell e Simon</i> <i>Connessionismo</i>

- **John R. Searle** ha proposto nel 1990 il **principio di connessione**, secondo il quale attribuire carattere inconscio a un fenomeno intenzionale è una contraddizione.
- Definiamo tre problemi principali della CS
 - a) *problema della relazione mente-corpo*
 - b) *problema dell'"homunculus"*, particolarmente evidente nella psicologia introspezionista, ma anche nel ricorso a fattori interni in certe correnti comportamentiste
 - c) *problema della possibilità di una CS autonoma.*
- La **CS classica** (CCS) ammette l'esistenza di stati mentali inconsci e la possibilità di macchine pensanti. Essa ha risposto al problema a) col *funzionalismo*, al

problema b) con l'*approccio computazionale*, al problema c) con la *teoria rappresentazionale degli stati mentali*.

- Il **funzionalismo** sostiene che gli stati mentali sono funzionali, nel senso che la loro essenza non è nel contenuto o nel substrato materiale, ma in quello che li ha causati e in quello che essi causano. Questa posizione non è incompatibile con la presunzione di esistenza di intelligenze non umane, né con la possibilità di implementazioni differenti degli stessi comportamenti intelligenti. Tra i sostenitori di questa visione il filosofo australiano David Malet Armstrong.
- Il **computazionalismo** sostiene che gli stati mentali sono computazionali, analoghi allo stato di un nastro di una TM. I processi mentali sono operazioni di calcolo che agiscono su simboli o sequenze di simboli. Il problema dell'*homunculus* è così risolto, giacché non esistono homunculi.
- Il **rappresentazionalismo** sostiene che gli stati mentali sono relazioni tra il soggetto e rappresentazioni mentali di eventi o processi. La CCS assume che le rappresentazioni sono simboli o sequenze di simboli correlati a eventi o processi fisici particolari. Per **Fodor** il **mentalese** (o **linguaggio del pensiero**) è il linguaggio che organizza le rappresentazioni mentali.
- Nell'ambito della CCS, per gli empiristi una macchina o un programma che implementi il mentalese deve poter interagire con l'ambiente esterno, e il modo di questa interazione è **reattivo**. Una tipica implementazione è il GPS, con le sue regole euristiche. Per i razionalisti una macchina o un programma che implementi il mentalese deve essere strutturata gerarchicamente in **moduli** deputati allo svolgimento di operazioni particolari, e il modo di questa interazione è **deliberativo**. I modelli razionalisti privilegiano la logica e le reti semantiche.
- Tra le critiche mosse alla CCS citiamo quella del filosofo **Thomas Nagel**, che nel 1974 scrisse ***What is it like to be a bat?*** Nagel sostiene che per un soggetto, apprendere simbolicamente una rappresentazione di uno stato mentale non equivale a provarlo soggettivamente.
- **John R. Searle** nel 1980 scrisse ***Mind, brains and programs***, dove descrive il famoso esperimento della stanza cinese. In questo attacco al computazionalismo, al rappresentazionalismo e all'AI forte, Searle sostiene che la mente è sempre associata a contenuti e a significati, che non potranno mai emergere da un qualunque tipo di rappresentazione simbolica o manipolazione sintattica. Il significato, l'intenzionalità, la coscienza si trovano cioè a un livello diverso da quello dei simboli e delle loro combinazioni. Certo il cervello è una macchina pensante, ma non è una TM.

IL PROBLEMA DELLA COSCIENZA

- Esistono molte forme di coscienza, come quelle associate con la visione, il pensiero, l'emozione, il dolore, ecc., l'**autocoscienza** (*self-consciousness*).
- Normalmente la posizione dei neuroscienziati è una delle seguenti:
 - a) la coscienza è un problema filosofico, quindi è meglio lasciarla ai filosofi

- b) la coscienza è un problema scientifico, ma è prematuro affrontarlo
- Il principale postulato dei neuroscienziati è l'esistenza di processi cerebrali temporalmente correlati con la coscienza.
- Secondo **F.Crick e C.Koch** (1995) la funzione biologica della coscienza visiva negli uomini è quella di produrre la migliore interpretazione della scena visiva alla luce dell'esperienza passata (sia in termini di memoria autobiografica che di memoria genetica) e di rendere questa interpretazione direttamente disponibile per un tempo sufficiente alle parti del cervello che devono scegliere una tra molte possibili risposte comportamentali.
- Anche **Ramachandran e Hirstein** (1997) rilevano l'utilità di una interpretazione cosciente della scena visiva, che in quanto *unica* elimina l'esitazione dei sistemi di risposta comportamentale.

VISIONE ARTIFICIALE

- Le ricerche e le metodologie proposte in questo ambito possono essere classificate in base al **livello di analisi** e lo **stadio di elaborazione**:
 - a) Livello computazionale. Specifica quali sono gli input visivi rilevanti e quali gli output, i vincoli matematici, ecc.
 - b) Livello algoritmico. Specifica gli algoritmi e le rappresentazioni
 - c) Livello implementativo.
 - d) Stadio di acquisizione dell'immagine
 - e) Stadio di analisi primitiva (bordi, dimensioni, distinzione figura-sfondo, ecc.)
 - f) Stadio di analisi dettagliata (modello 3D dell'oggetto, moto dell'oggetto, ecc.)
 - g) Stadio di interpretazione.
- **Stadio di acquisizione.** La digitalizzazione dell'immagine produce una matrice dei livelli di luminanza. A ogni pixel (x,y) è associato un livello di grigio, o **luminanza** $I(x,y)$ (invece, nelle immagini a colori, i livelli di rosso, verde, blu). Si applicano spesso filtri di pre-processing per il *noise-reduction* e lo *image-enhancement*, con algoritmi di convoluzione che usano maschere diverse. Tra gli operatori più importanti ricordiamo l'**operatore laplaciano**

$$\nabla^2 I = \partial^2 I / \partial x^2 + \partial^2 I / \partial y^2$$

e la **convoluzione con una Gaussiana**

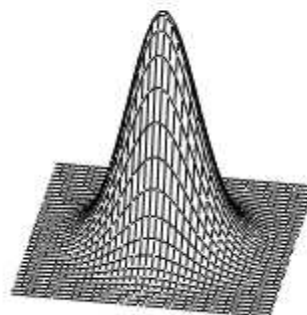
$$G(x,y) = (1/2\pi\sigma^2) \exp [-(x^2+y^2)/2\sigma^2]$$

$$\iint I(u,v) G(x-u,y-v) dx dy$$

L'applicazione su I del laplaciano evidenzia i contorni delle figure, precisamente nei *zero-crossing points* di $\nabla^2 I$.

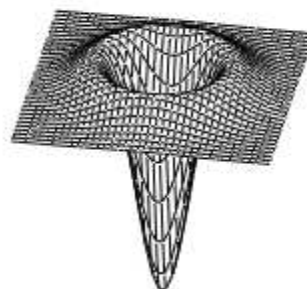
Esempio:

$$I(x,y) = \exp [-(x^2+y^2)/2\sigma^2]$$



$I(x,y)$

$$\nabla^2 I(x,y) = -(2/\sigma^2)[1 - (x^2+y^2)/2\sigma^2] \exp [-(x^2+y^2)/2\sigma^2]$$



$F(x,y)$

F

$$\nabla^2 I(x,y) = 0 \quad \Leftrightarrow \quad x^2 + y^2 = 2\sigma^2$$

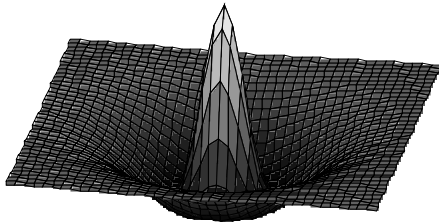
che è l'equazione di un cerchio centrato nell'origine.

Un difetto di questo operatore è che in caso di immagine rumorosa, l'immagine trasformata è ancora più rumorosa. La convoluzione con un filtro gaussiano attenua le componenti di alta frequenza spaziale, quindi tutti i rumori di alta frequenza.

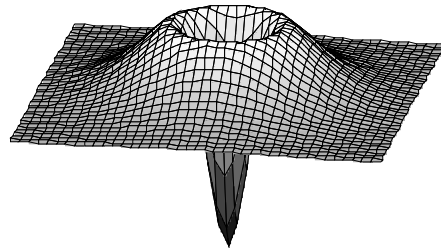
- **Stadio di analisi primitiva.** Uno dei problemi è determinare il contorno degli oggetti presenti nell'immagine. Ancora una volta si ricorre alla convoluzione. I primi tipi di maschere furono introdotti da Roberts nel 1965. **Marr** propose l'*operatore convoluzione col laplaciano della gaussiana* (o *operatore di Marr-Hildreth*), spesso abbreviato **LoG**. Marr definì *primal sketch* la trasformata $LoG I(x,y)$.

Come evidenziato da Marr e Hildreth, l'operatore *LoG* può essere approssimato dall'operatore **DoG** (*Differenza di gaussiane*)

$$f(x,y) = A \cdot e^{-a^2(x^2+y^2)} - B \cdot e^{-b^2(x^2+y^2)}$$



DOG



DOG

L'importanza dell'operatore DoG è che ha la stessa forma della curva ricavata da Robert W. Rodieck dai dati sperimentali sul campo recettivo delle cellule gangliari di tipo ON del sistema visivo di rana e di gatto. In quel contesto la formula utilizzata era

$$f(x,y) = [1 - (x^2 + y^2)/2\sigma^2] \exp[-(x^2 + y^2)/2\sigma^2]$$

Questo fatto ha indotto Marr a proporre l'operatore LoG, o l'equivalente DoG, come il più plausibile del punto di vista biologico. Gli inconvenienti sono l'amplificazione del rumore e talvolta l'emergenza di contorni inesistenti.

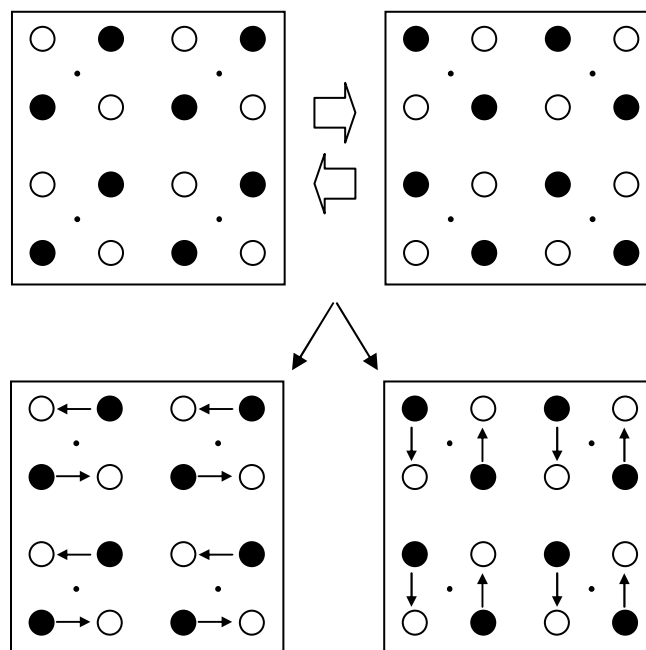
Si noti che $f(x,y) = -\pi\sigma^4 \nabla^2 G(x,y)$

- Attualmente si utilizza il più moderno algoritmo sviluppato da **John Canny** nel 1983. L'elaborazione è basata su un operatore che combina la gaussiana con altri semplici operatori capaci di rilevare i gradienti di grigio nelle direzioni orizzontale e verticale.
- Un altro problema dello stadio di analisi primitiva è la separazione figura-sfondo e l'analisi della forma degli oggetti.
- Def. **momenti geometrici**

$$m_{pq} = \sum_i \sum_j x_i^p y_j^q I(x_i, y_j) \quad p, q \geq 0$$

I momenti geometrici di un'immagine possono caratterizzare alcuni aspetti, e si possono rilevare delle invarianze rispetto a certe trasformazioni (rotazione, centrimento, standardizzazione delle dimensioni, ecc.)

- Def. L'aggettivo **globale** (*global*) si riferisce ad ampie regioni del campo visivo, in opposizione a **locale** (*local*), che si riferisce a regioni molto piccole, dell'ordine dell'ampiezza dei campi recettivi delle cellule della regione parafoveale dell'area *V1* ($\approx 1^\circ$) o *V4* ($\approx 5^\circ$).
- Def. **Problema di corrispondenza** (*Correspondence Problem*). Data una successione di immagini di ripresa di un oggetto in movimento, determinare quali caratteristiche di una data immagine sono associate a quali caratteristiche dell'immagine seguente.
- **Ullman** (1979) propose un algoritmo che risolve il problema di corrispondenza localmente. Questo algoritmo calcola tutti gli accoppiamenti possibili e iterativamente converge alla corrispondenza che minimizza la distanza. Grzywacz e Yuille (1986) hanno implementato un algoritmo simile con una ANN.
- Ma il sistema visivo utilizza probabilmente anche informazioni globali per risolvere il problema di corrispondenza, come dimostra l'esperimento dei **quartetti bistabili** (*Ramachandran e Anstis*, 1983). Su uno schermo sono presentati in successione due fotogrammi (quadrati superiori in figura). Il rendimento percettivo è quello di punti luminosi (cerchi bianchi) che oscillano tutti in senso orizzontale o verticale (quadrati inferiori), mai una combinazione mista di movimenti verticali e orizzontali. Qui il vincolo globale prevale su quello locale nel determinare la soluzione al problema di corrispondenza.



- Def. Il **problema di segmentazione** (*Segmentation Problem*). Data un'immagine, determinare quali caratteristiche, in punti diversi del campo visivo, sono parti di uno stesso oggetto. Questo problema è anche noto come problema di unione (*binding problem*).

- **Stadio di analisi dettagliata.** Tra i problemi implicati la determinazione della forma 3D da una o più immagini 2D; la determinazione del moto di un oggetto nel campo visivo o del moto proprio (*ego-motion*); estrazione di un modello stilizzato 3D (*scheletro*). Gli algoritmi si dividono in *strutturali* e *statistici*. Per quanto riguarda la determinazione della forma 3D secondo la teoria di Marr e Poggio (1979) essa dipende sostanzialmente dalla differenza nelle immagini sinistra e destra di una ripresa stereoscopica (*primal sketches*). L'algoritmo di confronto assegna un valore di *disparità binoculare* a ciascuna coppia di pixel corrispondenti delle due immagini. In realtà il pattern risultante è compatibile con più ricostruzioni 3D, sicché i due autori hanno aggiunto due vincoli ulteriori:
 - a) **Unicità:** ogni pixel di ciascuno dei due *primal sketches* è la proiezione di un solo punto della ricostruzione 3D
 - b) **Continuità:** la funzione di disparità è continua in x e in y .
 La ricostruzione così ottenuta è detta **$2\frac{1}{2}$ - dimensional sketch**, da sottoporre a un ulteriore stadio di elaborazione, mai descritto completamente dagli AA. Si tratta di utilizzare combinazioni di oggetti 3D come cilindri e poliedri.
- **Stadio di interpretazione.** Riconoscere o identificare un oggetto implica che il soggetto ne abbia una precedente rappresentazione. Distinguiamo l'*approccio basato sull'oggetto* e l'*approccio basato su più immagini 2D*. L'**approccio basato sull'oggetto** è sostenuto da Marr: la forma ricostruita 3D dell'oggetto è indipendentemente dal punto di osservazione. Il processo di riconoscimento è gerarchico: parte dalla forma globale e procede per successive segmentazioni delle parti costituenti mediante cilindri approssimanti. Recentemente sono stati affinati gli algoritmi di segmentazione e si sono aggiunte ai cilindri altre forme approssimanti, come poliedri (Waltz, 1975), sfere (Badler e Bajcsy, 1978), superquadratiche (Pentland, 1986), *geoni* (Biederman, 1987).
- La **teoria del confronto di sagome (template matching theory)** assume che nella mente ci sia una qualche rappresentazione o sagoma per ciascuna delle configurazioni riconoscibili. La configurazione in entrata viene confrontata con tutte le sagome possibili, per trovare quella che meglio si adatta. Ma, considerando il caso di una "A", per spiegare il riconoscimento della stessa lettera sottoposta a rotazioni, modificazioni di scala, modificazioni tipografiche, ecc. occorrerebbe aver memorizzato una quantità immensa di sagome, alcune delle quali mai viste prima!
- La **teoria dei prototipi** assume che la rappresentazione dell'oggetto in memoria è piuttosto una sintesi statistica di tutte le precedenti esperienze
- La **teoria dell'analisi delle caratteristiche (feature analysis)** assume che il riconoscimento di un oggetto si attui sulle sue caratteristiche invarianti rispetto alla traslazione, rotazione, inversione, ecc. Queste teorie furono molto diffuse negli anni '60 e '70 (Selfridge 1959; Neisser, 1967; Lindsay, Norman 1972). I dati neurofisiologici di Hubel e Wiesel tendono a confermare queste teorie, criticabili per altri aspetti:
 - i) Scarsa importanza al ruolo svolto dal contesto e dalle aspettative

- ii) Non considerano le relazioni tra le caratteristiche semplici
 - iii) Sono difficilmente estensibili a stimoli complessi, come volti o edifici
 - iv) Spesso sembra che il riconoscimento sia basato sulla forma globale gestaltica piuttosto che sulle singole caratteristiche
- Le **teorie strutturali** (tra cui quella di Winston, 1977) si fondano più sulle relazioni tra gli elementi che non sull'insieme degli elementi presenti nell'immagine (p.es. un triangolo è definito dalla relazione tra i lati più che dalla presenza di tre segmenti).
 - La **teoria computazionale di Marr** (1982) è una teoria cognitiva che considera la visione, tanto del cervello quanto della macchina, una costruzione di descrizioni-rappresentazioni simboliche via via più dettagliate e ricche di informazioni. Le variazioni di luminanza della scena definiscono i contorni di un abbozzo primario, bidimensionale (*primal sketch*). La rappresentazione successiva è l'abbozzo 2 ½-D (*2 ½-Dimensional sketch*) che in parte esplicita la tridimensionalità dell'oggetto, basandosi sui dati forniti dalla disparità binoculare, dal movimento, dalle ombre. La costruzione della rappresentazione finale 3-D, tipicamente formata da cilindri, si basa su un *modello basilare di simmetria* che traspone i contorni dell'abbozzo 2 ½-D in assi di simmetria. Questo passaggio, a differenza degli altri due, è presumibilmente di tipo top-down, ed è centrato sull'oggetto, mentre le altre due rappresentazioni erano centrate sul punto di vista dell'osservatore.
 - La **teoria del riconoscimento per componenti** di **Biederman** (*Theory of Recognition by Components, RBC*, 1985, 1987, 1990) postula 36 solidi primitivi detti **geoni** (*ioni geometrici*). Il riconoscimento nasce dal confronto delle disposizioni spaziali dei geoni con dati analoghi presenti in memoria. Biederman ha utilizzato il modello *RBC* per interpretare i rendimenti di soggetti umani in test di riconoscimento visivo.
 - Altre teorie sfruttano le differenze di ombreggiatura, lo stato di moto e la disponibilità di più immagini bidimensionali da punti di osservazione diversi. Tutte queste teorie presuppongono una memoria a lungo termine già caricata.
 - L'**approccio basato su più immagini 2D** poggiano sugli **invarianti** e/o sulle **caratteristiche distribuite**. Gli invarianti sono quantità globali che non cambiano sotto certe trasformazioni. I pionieri in questo campo furono **Duda e Harth** (1973). Le caratteristiche distribuite hanno rappresentazione subsimbolica.

RAPPRESENTAZIONE DELLA CONOSCENZA

- La conoscenza che si immette nei programmi di AI è conoscenza umana, quindi il modo di rappresentarla deve rispecchiare un modello di codifica della conoscenza nel cervello o nella mente.
- Le rappresentazioni simboliche classiche proposte dall'AI e dalla CCS sono di tipo **dichiarativo** (*database, a rete, frames, scripts*) o **procedurale** (*sistemi di produzioni*).

- Il **database** è un insieme strutturato o meno di caselle contenenti una quantità di informazione.
- I **sistemi basati sui linguaggi della Logica formale**, composti da assiomi e regole di inferenza, sono efficaci solo se
 - a) Il linguaggio è abbastanza flessibile e potente da esprimere in modo semplice tutte le proprietà dei fatti e delle relazioni validi nel dominio di conoscenza. Questa condizione purtroppo non è quasi mai soddisfatta.
 - b) Esiste una strategia automatica di deduzione che permette in tempi ragionevoli di verificare una proposizione. A questo scopo **John Alan Robinson** ha ideato il **principio di risoluzione**, che è stato incorporato nel linguaggio **PROLOG** (inventato da Alain Cormerauer e Phillip Roussel nel 1972). La **programmazione logica** fatta col PROLOG si riduce ad editare un elenco di fatti e di regole. Una volta inserita una domanda il compilatore del linguaggio attiva il principio di risoluzione e una strategia euristica di ricerca in profondità per fornire una risposta, che equivale alla dimostrazione di un teorema.

Il **Calcolo dei Predicati del Primo Ordine** è stato sviluppato per potenziare le capacità espressive del Calcolo Proposizionale. L'elemento centrale è il predicato, che è una funzione di uno o più argomenti. P.es.

$$\forall x (if\ cane(x)\ then\ abbaia(x))$$

che si può leggere “tutti i cani abbaiano”. cane(x) e abbaia(x) sono predicati. L'attributo “del primo ordine” si riferisce al fatto che i quantificatori possono agire solo su singole variabili.

Il **Calcolo dei Predicati del Secondo Ordine** permette di applicare i quantificatori anche ai predicati, sicché è possibile p.es. definire la relazione di uguaglianza con una proposizione del tipo

$$\forall P ((x=y)\ iff\ (P(x)\ iff\ P(y)))$$

Nella terminologia di Tarsky x sarebbe una variabile individuale, P una variabile predicativa.

Mentre il principio di risoluzione è generalizzabile al **Calcolo dei Predicati del Primo Ordine**, non si conoscono ancora strategie automatiche di deduzione per il **Calcolo dei Predicati del secondo Ordine**. Oggi il **Calcolo dei Predicati del Primo Ordine** è largamente usato per costruire i programmi di AI.

- Le rappresentazioni dichiarative **a rete** della conoscenza usano:
 - a) reti semantiche
 - b) grafi concettuali
 - c) reti Pathfinder

- Le **reti semantiche** sono formate da *nodi* (i concetti) uniti da *archi* labellati (le relazioni). Furono introdotte da **Collins e M.R.Quillian** (1968). Possono essere usate in **modo statico** e **dinamico**. Nel modo dinamico agli archi è associata una funzione di **attivazione** che si propaga in modo automatico o controllato tra i nodi della rete. Una sufficiente attivazione rende disponibili questi elementi nella memoria di lavoro. Nel modello di rete semantica di Collins e Loftus il collegamento tra due nodi ha lunghezza inversamente proporzionale al tempo di propagazione dell'attivazione da un nodo all'altro.

Nei **grafi concettuali**, introdotti da John F.Sowa (1983), i nodi rappresentano sia concetti che relazioni (permettono di inserire relazioni tra molti concetti), mentre gli archi non sono labellati, ma talvolta vi si associa un peso. Come anche le reti semantiche, hanno la stessa potenza del *Calcolo dei Predicati del Primo Ordine*.

Le **reti Pathfinder** hanno pesi associati agli archi, ma non assegnate a priori, bensì determinati da un algoritmo di **analisi di prossimità**, essendo predefinita una funzione di prossimità tra concetti (Roger W.Schvaneveldt).

I **frames** furono ideati da **M. Minsky** nel 1975, sulla base dei concetti di **schema** di F.C.Barlett e di **paradigma** di Th.S.Kuhn. Un *frame* è una struttura-dati presente in memoria, per Hofstadter (1979) una “*rappresentazione algoritmica di un contesto*”, una collezione di nozioni implicite con cui vengono rappresentate situazioni o eventi stereotipati (p.es. una tipica festa di compleanno, un tipico soggiorno). Un frame è formato da livelli superiori, che rappresentano gli aspetti invarianti (p.es. ogni soggiorno è una stanza con soffitto), e livelli terminali (**slots**) con valori assegnati implicitamente (**default values**) in assenza di specifici dati dell'esperienza. I *valori di default* possono essere sostituiti da altri frames, man mano che l'esperienza rivela i dettagli della situazione specifica. I frames hanno ispirato la creazione di molti linguaggi di programmazione il cui capostipite è stato **KRL** (Daniel Bobrow e Terry Winograd, 1977). Il limite dei *frames* è che non c'è modo di impedire la dichiarazione di relazioni contraddittorie in quanto prescindono dal livello semantico.

R.C. Schank e R.P Abelson (1977) denominano **scripts** una sequenza stereotipata di eventi o di azioni che caratterizzano un certo contesto. Esempi di *script* sono le azioni da compiere quando si entra in un negozio o si deve prendere il treno. Il **MOP (Memory Organization Packet)** è una rielaborazione più recente e più flessibile dello *script* (Schank, 1982). Un *MOP* è un gruppo organizzato di *scene*.

I **Sistemi di Produzioni**, introdotti da **Emil Leon Post** negli anni '20 sono formati da entità elementari dette produzioni, nella forma generale:

se (condizione) allora (azione)

L'utilizzo di un sistema simile richiede un sistema di controllo capace di monitorare l'ambiente e di verificare quali sono le condizioni valide in un dato istante, nonché aggiornare lo stato dell'ambiente in base alle azioni effettuate.

- Tra i lavori più ambiziosi **CYC**, iniziato da **Lenat** (1984), capace di estrarre, per mezzo di inferenze, conoscenza implicita da una base di conoscenze di tipo dichiarativo.

ARCHITETTURE COGNITIVE

- *Def.* Un'architettura cognitiva è un sistema che gestisce le risorse cognitive di un essere (naturale o artificiale) con capacità di percezione e azione in un ambiente.
- Newell propose l'architettura **SOAR** (**State, Operator And Results**), basato sul **Principio della Razionalità Limitata**, proposto da Simon nel 1957. Secondo questo principio l'agente deve comportarsi in modo non necessariamente ottimale per il raggiungimento dei suoi scopi, limitatamente alle risorse disponibili. Da questo principio Newell ha ricavato nel 1982 l'Ipotesi della Massima Razionalità. Secondo il Principio dello Spazio Problematico l'attività dell'agente sarà descritta da:
 - a) un insieme di stati di conoscenza
 - b) un insieme di operatori per passare da uno stato all'altro
 - c) dei vincoli all'applicazione degli operatori
 - d) un sistema di controllo basato su conoscenze che decide quali operatori applicare

SOAR è formato dai seguenti componenti:

- 1) una memoria a lungo termine
- 2) una memoria di lavoro
- 3) il sistema di elaborazione-decisione
- 4) il sistema percettivo-motorio

Il pregio e il limite di **SOAR** è la sua capacità di selezionare conoscenze da un'ampia base di conoscenze prememorizzate.

- Nel 1993 Anderson propose l'architettura **ACT-R** (**Adaptive Control of Thought**), basato sul **Principio della Razionalità**. Secondo questo principio, l'agente deve comportarsi sempre in modo da ottimizzare il suo adattamento all'ambiente.

Secondo Anderson occorre invece focalizzarsi sulla natura del compito. A differenza di **SOAR**, **ACT-R** ha anche una rappresentazione subsimbolica della conoscenza. Il livello simbolico è formato da **chunks** (parte dichiarativa), simili ai *frames* di Minsky, e da regole di produzione (parte procedurale). Ogni *chunk* presente a un dato momento nella memoria di lavoro ha associata un'**attivazione teorica**

$$A_i = B_i + \sum_j W_j S_{ji}$$

Dove W_j è la *j*-esima **sorgente di attivazione** del *chunk* *i*-esimo (normalmente gli altri *chunks* contenuti negli *slots* del *chunk* *i*-esimo), mentre S_{ji} è la **forza di**

associazione tra W_j e il *chunk* i -esimo. S_{ji} e B_i sono modificabili dall'esperienza, con valori iniziali

$$B_i^0 = \log P_i / (1 - P_i)$$

$$S_{ji}^0 = \log m - \log n$$

con m numero totale dei *chunks* presenti in memoria, n numero di *chunks* con uno *slot* contenente proprio W_j come default. Una conseguenza è che maggiore è n , minore l'attivazione. In questo modo il modello spiegherebbe il ***fan-effect***, evidenziato sperimentalmente da Anderson nel 1974 (nei test di recupero guidati da indizi, maggiore è il numero di item cui era stato associato l'indizio in questione, peggiore è la performance).

L'attivazione effettiva è

$$A_i' = A_i + N(\sigma) - P_m M$$

con $N(\sigma)$ rumore gaussiano a media nulla e varianza σ^2 , $M \in [0, 1]$ il ***mismatch*** ammissibile nel confronto tra il *chunk* i -esimo e la (*condizione*) di una generica produzione.

Ad ogni passo viene considerata l'attivazione di tutti i *chunks* presenti nella (*condizione*) di ciascuna produzione; se l'attivazione supera una certa soglia per tutti i *chunks* di una data (*condizione*), la corrispondente azione viene eseguita.

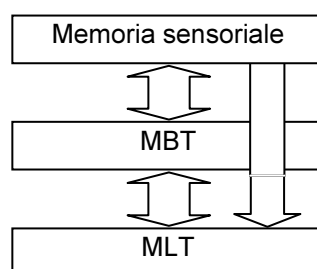
LA PSICOLOGIA COGNITIVA

- La CS ha le sue origini nell' AI classica e nella psicologia cognitiva.
- Lo status attuale della ***psicologia cognitiva*** è assai frammentario, comprendendo coloro che sperimentano in ristretti ambiti senza avvertire il bisogno di avanzare modelli teorici, coloro che avvertono questa esigenza in domini specifici (alcuni di questi si ispirano al HIP), coloro che cercano modelli teorici più generali.
- Lo ***HIP (Human Information Processing)*** possiede, secondo Palmer e Kimichi (1986),
 - a) Una descrizione informazionale (informazione intesa in senso semantico)
 - b) La scomponibilità ricorsiva (ogni processo è formato da sottoprocessi)
 - c) La continuità del flusso informazionale (dal passato verso il futuro)
 - d) La fisicità dell'implementazione

LA MEMORIA

- Distinguiamo ***test di natura episodica*** (p.es. imparare una specifica lista di item) e ***test di natura semantica*** (p.e. rispondere vero/falso a un quesito di cultura generale).

- Distinguiamo **risposte di richiamo** (il soggetto deve riferire l'item appreso) e **risposte di confronto** (il soggetto deve fornire un giudizio di confronto tra lo stimolo di test e quello appreso).
- **Decadimento esponenziale**. In una prova di richiamo la prestazione decresce esponenzialmente col tempo di ritenzione tra presentazione e richiamo.
- **Saturazione**. In una prova di richiamo di sequenze la prestazione è sempre inferiore al 100% se il numero di lettere/cifre è superiore a 7 ± 2
- In una serie di prove seriali, la percentuale di risposte corrette in funzione dell'indice dell'ordine di presentazione dell'item ha un andamento a U, sicché si ricordano meglio gli items che occupano i primi posti (**effetto primacy**) e gli ultimi posti (**effetto recency**) della lista.
- **Interferenza**. Apprendimenti successivi interferiscono. Se l'apprendimento successivo peggiora il recupero dell'item appreso in precedenza si ha **interferenza retroattiva**, se invece peggiora il recupero dell'item nuovo si ha **interferenza proattiva**.
- **Principio di specificità della codifica**. In un test episodico guidato da indizi, la prestazione migliora se l'indizio usato è lo stesso di quello che era presente al momento della memorizzazione
- Distinguiamo **modelli strutturali** e **modelli funzionali** della memoria.
- Distinguiamo **modelli gerarchici** e **modelli non gerarchici** della memoria.
- I modelli cognitivisti di funzionamento della memoria umana si dividono grosso modo in due gruppi, quelli sviluppati negli anni '60 e quelli sviluppati negli anni '70.
- I modelli del primo gruppo, che cercavano di spiegare le **curve di oblio** in funzione dell'**intervallo di ritenzione** (precedente la fase di richiamo), furono quelli basati sulla **teoria dell'interferenza** (l'oblio è dovuto all'interferenza tra gli items, indipendentemente dall'intervallo di ritenzione) e quelli basati sulla **teoria del doppio magazzino** (transizione dell'informazione da MBT a MLT già dopo 30 sec di ritenzione).
- Nel modello strutturale di **Atkinson e Shrifin** (1968) sono contemplati tre processi in serie
 - 1) un **registro sensoriale** (RS) per ciascuna modalità sensoriale, che trattiene l'informazione per tempi brevissimi (da 250 ms a 1 s per il canale visivo, 3-4 s per il canale uditivo) e ha una capacità illimitata
 - 2) una **memoria a breve termine** (MBT) a capacità limitata, capace di trattenere 7 ± 2 elementi (chunks) per $\approx 10-20$ sec.
 - 3) una **memoria a lungo termine** (MLT), con capacità e durata illimitate



Il modello specifica la probabilità di ricordo esatto dell'item a tra n items dopo un intervallo di ritenzione di k sec.

$$P(k) = \sum_i P(k|i) P_{MBT}(i)$$

dove $P(k|i)$ è la prob. che dopo k sec. a sia nella MLT avendo transitato i sec. nella MBT, mentre $P_{MBT}(i)$ è la prob. che a sia in MBT dopo i sec.

$$P(k|i) = 1 - (1-g)\exp(-i\theta\tau^{k-1}) \Psi(k-i)$$

g è la prob. di indovinare la risposta corretta ($\approx 1/n$)

θ è la frazione di informazione che viene trasferita dalla MBT alla MLT

τ è il tasso di decadimento spontaneo della MLT

La stima di $P_{MBT}(i)$ si basa sull'ipotesi che la MBT sia strutturata in r caselle di informazione con probabilità di cancellazione spontanea d . A un dato istante l'informazione può trovare posto nella MBT solo se ci sono caselle disponibili. Se d tende a zero e b è la prob. di transizione dal registro sensoriale alla MBT si ricava

$$P_{MBT}(i) = (1-b) \quad \text{se } i=0$$

$$P_{MBT}(i) = b(1-b/r)^{i-1} (b/r) \quad \text{se } i>0$$

Il meccanismo proposto tiene conto dell'interferenza, ma in modo semplicistico, e trascura la MLT.

- Il **modello dei livelli di elaborazione** di **Craik e Lockart** (1972) considera la memoria un sistema unitario, cioè nega l'esistenza di più sottosistemi (modelli *plurimemoria*). La durata delle tracce dipende dal tipo di analisi effettuata sullo stimolo (sensoriale, semantica, ecc.) . Esisterebbero differenti livelli di elaborazione, essendo la "profondità" di elaborazione proporzionale al numero di connessioni semantiche che un item è in grado di formare con gli altri in memoria.
- **Tulvin** (1972) propose una distinzione della MLT in **memoria episodica**, un magazzino di eventi autobiografici collocati in uno spazio e in un tempo precisi, e **memoria semantica**, contenente significati, concetti, conoscenze. In seguito (1985) Tulvin contemplò anche una **memoria procedurale**, contenente abilità percettive-cognitive-motorie per lo più automatiche.
- Sul problema se esista o meno, oltre a una memoria **proposizionale**, anche una memoria fotografica (o **analogica**), è incentrato il lavoro del canadese **Paivio**

(1975). Egli avanzò la **teoria del doppio codice** (parole che richiamano immagini sono richiamate più prontamente di parole che non evocano immagini). Paivio sosteneva l'esistenza di due centri separati, uno di elaborazione pittorica e figurale, l'altro linguistica. Le unità figurali sarebbero organizzate in modo sincrono e simultaneo, mentre le unità linguistiche in modo sequenziale.

- Le **reti semantiche** sono formate da *nodi* (i concetti) uniti da *archi* labellati (le relazioni). Possono essere usate in **modo statico** e **dinamico**. Nel modo dinamico ai nodi è associato uno stato di **attivazione** che si propaga in modo automatico o controllato ad altri nodi attraverso gli archi. Una sufficiente attivazione rende disponibili questi elementi nella memoria di lavoro.
- **Collins e Quillian** (1969) proposero per la MLT un modello di **memoria semantica** strutturato ad albero, in cui ogni nodo rappresenta una categoria (di oggetti, animali, ecc.) che appartiene alla categoria rappresentata dal nodo padre. Inoltre, a ogni nodo sono associate delle proprietà particolari. Per verificare la realtà psicologica del modello gli autori sottoponevano a volontari frasi come “*uno squalo è un pesce*”, “*uno struzzo è un pesce*”, ecc. alle quali esso dovevano rispondere *vero* o *falso*. Si osservò che effettivamente la latenza misurata era proporzionale alla distanza tra le categorie nell'albero semantico (**category size effect**).
- Nel modello di **Collins e Loftus** (1975) esistono due reti semantiche collegate: una **rete lessicale** (i nodi sono nomi e gli archi relazioni lessicali) e una **rete concettuale** (i nodi sono concetti e gli archi relazioni concettuali). il collegamento tra due nodi ha lunghezza inversamente proporzionale al tempo di propagazione dell'attivazione da un nodo all'altro.
- Il **modello ACT** di **Anderson** (1976) rappresenta conoscenze **dichiarative** (fatti, concetti, opinioni, ecc.) e **procedurali** (sequenze di azioni finalizzate). La conoscenza dichiarativa è strutturata come nel modello di Collins e Loftus, ma gli archi (del tipo **soggetto-predicato** o **relazione-argomento**) hanno un coefficiente di **forza** dell'associazione. La conoscenza procedurale è composta di **produzioni**, del tipo **condizione-azione**. Il modello è troppo complicato per poterne valutare le reali potenzialità.
- La **teoria degli insiemi** di Meyer (1970) e il **modello del confronto dei tratti** di Smith, Shoben e Rips (1974) non considerano la MLT come rete semantica. I concetti non sono nodi ma piuttosto insiemi di attributi.
- Il modello di Hintzman (1986).
- Il **modello di Tulving e Thomson** (1963) (**encoding specificity theory**) afferma che la probabilità che un'informazione venga recuperata dipende solo dal grado di somiglianza tra le caratteristiche presenti nell'ambiente al momento del recupero e le caratteristiche codificate nella traccia mnemonica.
- La teoria di Tulving si contrappone alle cosiddette **teorie a due processi** (**two process theory**), che ipotizzano un processo di generazione dei possibili candidati

dell'item ricercato, e un processo di riconoscimento, che seleziona l'item, in modo da spiegare alcuni effetti:

- i) il richiamo libero è meno facile del richiamo guidato, che è meno facile del riconoscimento
- ii) le parole con più alta frequenza nel linguaggio parlato sono più facili da recuperare rispetto a quelle a più bassa frequenza, mentre nel riconoscimento avviene il contrario.

- I **modelli funzionali di memoria** considerano l'item immagazzinato sotto forma di vettore, dove ogni componente è associata a una caratteristica dell'item, con valore numerico funzione del grado di presenza.

Un limite è dato dalla fissità del numero di caratteristiche (la dimensionalità dei vettori).

- Nel 1980 **Raajmakers e Shiffrin** proposero il modello **SAM (Search of Associative Memory)**.

Il modello prevede una MBT, capace di contenere temporaneamente contesti e items.

La generica (*j-esima*) unità di informazione immagazzinata nella MLT è detta **immagine (memory image)**, formata da:

$V^{(j)}$ vettore che codifica l'item *j-esimo*

$F^{(j)}$ vettore la cui componente *i-esima* è la forza di associazione tra gli item *j-esimo* e *i-esimo*.

$C^{(j)}$ vettore la cui componente *k-esima* è la forza di associazione tra l'item *j-esimo* e il contesto *k-esimo*.

$F^{(j)}$ e $C^{(j)}$ sono funzioni di *t* durante il tempo di permanenza di $V^{(j)}$ nella MBT. In particolare se la MBT contiene l'item *j-esimo* in $(t, t+\Delta t)$

$$F_j^{(j)}(t+\Delta t) = F_j^{(j)}(t) + (a\Delta t + d) [1 + \Psi(h - 1/3) + \Psi(h - 2/3)]/2$$

Se la MBT contiene l'item *i-esimo* in $(t, t+\Delta t)$

$$F_i^{(j)}(t+\Delta t) = F_i^{(j)}(t) + (b\Delta t + d) [1 + \Psi(h - 1/3) + \Psi(h - 2/3)]/2$$

Se la MBT contiene anche il contesto *k-esimo* in $(t, t+\Delta t)$

$$C_k^{(j)}(t+\Delta t) = C_k^{(j)}(t) + (c\Delta t + d) [1 + \Psi(h - 1/3) + \Psi(h - 2/3)]/2$$

Gli incrementi effettivi sono corretti in senso probabilistico (*h* è una v.a. distribuita uniformemente in (0,1)).

Richiamo. L'indizio di richiamo è dato da un item *V* e da un contesto. Per ciascuna

immagine presente in memoria si calcola la forza di associazione con l'item di

richiamo (F), e la forza di associazione col contesto di richiamo (C).

L'immagine

richiamata è quella con massimo prodotto FC .

Riconoscimento. Il pattern di richiamo (sia il j -esimo item) è riconosciuto (come item già memorizzato) solo se

$$FA \equiv \sum_k (F_k^{(j)})^r (C_k^{(j)})^s \geq thr.$$

- Nel 1984 **Hintzman** propose il modello strutturale **MINERVA2**.

Le componenti di $V^{(j)}$ possono valere

$v_i = -1$ caratteristica assente

$v_i = +1$ caratteristica presente

$v_i = 0$ mancanza di informazioni a riguardo

Se v_i è il valore teorico da assegnare alla componente i -esima dell'item $V^{(j)}$, il valore assegnato è $v_i \Psi(x-L)$, con x v.a. distribuita uniformemente in $(0,1)$, L parametro in $[0,1]$. Questa correzione rende conto dell'oblio spontaneo della MLT.

Richiamo. L'indizio di richiamo V genera per ogni item presente in memoria $V^{(j)}$, una **similarità**

$$S^{(j)} = (1/N_R) V \cdot V^{(j)}$$

dove N_R è il numero di componenti non nulle sia in V che in $V^{(j)}$.

La i -esima caratteristica dell'item richiamato dalla memoria vale quindi

$$R_i = \text{sign}[\sum_j (S^{(j)})^3 V_i^{(j)}]$$

Riconoscimento. Il pattern di richiamo V è riconosciuto solo se

$$F \equiv \sum_j (S^{(j)})^3 \geq thr. \quad (\text{intensità dell'eco})$$

Il modello di Hintzman spiega il fenomeno osservato su soggetti umani per cui mentre in prove di richiamo si osserva sia **list-length effect** (riduzione delle performance all'aumentare della lunghezza della lista degli items) che **list-strength effect** (miglioramento delle performance all'aumentare del numero di ripetizioni dell'item nella lista), nelle prove di riconoscimento, si osserva solamente il primo.

- Tra i cosiddetti **modelli a matrice della memoria** citiamo il **Brain-State-in-a-Box (BSB)**, proposto da Anderson, Silverstein, Ritz e Jones nel 1977. La memoria è rappresentata dalla matrice con elementi A_{ij} che rappresentano la forza di associazione tra gli items i -esimo e j -esimo.
- I **modelli vettoriali** (o **olografici**) della memoria si basano sull'analogia con il processo di formazione e lettura degli ologrammi ottici.

Un **ologramma** contiene in una superficie 2D tutta l'informazione luminosa di un oggetto 3D. Mentre una fotografia cattura solo la luce riflessa dall'oggetto in una direzione, l'ologramma sfrutta gli effetti di **interferenza** (costruzione) e **diffrazione** (lettura). In fase di costruzione una sorgente emette un **raggio di riferimento** e un raggio che colpisce l'oggetto e ne viene riflesso. Il **raggio riflesso** interferisce col raggio di riferimento formando una figura di interferenza in corrispondenza di una lastra fotografica. Spostando la sorgente le figure di interferenza risultanti si sovrapporranno alla prima. In fase di lettura un raggio di riferimento viene fatto incidere sull'ologramma, la luce è diffratta e forma a una certa distanza una figura di diffrazione. Aggiungendo altri raggi di riferimento si aggiungono via via altre figure di diffrazione a formare l'immagine 3D originaria.

Lo stesso inventore dell'olografia, Dennis Gabor, propose nel 1968 un'analogia tra ologramma e traccia di memoria distribuita. Il raggio di riferimento sarebbe analogo al contesto e il raggio riflesso all'item.

Nel 1973 A.Borsellino e T.Poggio hanno dimostrato che i processi ottici sopradescritti trovano la loro generalizzazione nelle algebre di convoluzione e correlazione. La convoluzione è l'analogo dell'interferenza mentre la correlazione è l'analogo della diffrazione.

- Tutti i modelli descritti finora sono modelli associazionistici. I modelli connessionistici in teoria dovrebbero invece descrivere i processi di memoria come fenomeni emergenti da una collettività di entità microscopiche, ma ciò non si è ancora verificato, sicché in pratica modelli connessionistici e associazionistici coincidono
- I modelli connessionistici più semplici, come quelli descritti, sono affetti da **interferenza catastrofica**, cioè l'informazione memorizzata è degradata dalle nuove informazioni (una sorta di **interferenza retroattiva**)

PERCEZIONE E ATTENZIONE

- La percezione ha natura tipicamente ricostruttiva nei confronti del pattern di stimolazione
- Stimolazioni sufficientemente intense producono effetti postumi transitori (nel caso della percezione visiva noti come **post-immagini** o **immagini postume**) Particolari pattern producono percezioni distorte (**illusioni**). Un esempio è la percezione di contorni illusori (p.es. il triangolo di Kanizsa)
- Particolari pattern statici producono cambiamenti percettivi (nel caso della percezione visiva sono le **figure ambigue**)
- Tipicamente percepiamo visivamente con maggiore efficacia oggetti considerati più plausibili o probabili (effetto della probabilità)
- Le risorse attentive sono limitate
- Anche in assenza di attenzione è possibile che il soggetto memorizzi uno stimolo per lui significativo

- I tempi di reazione nei compiti di ricerca visiva sono indipendenti dal numero complessivo degli elementi presenti quando lo stimolo target differisce dagli stimoli distrattori per una sola caratteristica; in caso contrario sono proporzionali al numero delle caratteristiche
- E' larga convinzione che i processi percettivi siano un fenomeno in gran parte bottom-up, una sorta di sofisticato filtro automatico, indipendenti da altre attività cognitive, in cui l'attenzione avrebbe un ruolo di ottimizzatore.
- Possiamo distinguere le teorie sulla percezione in base al ruolo assegnato allo stimolo fisico e alla presenza o meno di più stadi di elaborazione.
- **Teorie basate sullo stimolo.** Il rendimento percettivo dipenderebbe esclusivamente dalle caratteristiche dello stimolo. Citiamo le teorie a stadi di *Marr e Poggio* e *Pandemonium*, e le teorie olistiche di *Gibson*, *del confronto di sagome*, *dei prototipi*, e *RBC*
- **Teorie basate sui fattori interni** (o *costruttiviste*). Il rendimento percettivo dipenderebbe anche da altri fattori interni (emozioni, conoscenze, ecc.). Citiamo le teorie a stadi di *Rock*, *dell'informazione strutturale e connessionistiche*, e la teoria olistica della *Gestalt*.
- **Teoria computazionale di Marr e Poggio.** Riguarda esclusivamente la percezione visiva ed è stata esposta nei capitoli precedenti
- **Teoria Pandemonium** (Selfridge e Neisser, 1959). È una rete multistrato *feed-forward*, con unità dette **demoni**. Gli strati sono, nell'ordine,
 - Demoni dello stimolo** – rivelano la distribuzione della stimolazione
 - Demoni dei tratti** – rivelano le caratteristiche
 - Demoni cognitivi** – rivelano configurazioni di caratteristiche
 - Demoni decisionali** – categorizzazione dello stimolo
 Ad esempio, nel caso di riconoscimento di una lettera dell'alfabeto (A,B,..) i demoni dei tratti riconoscono segmenti orizzontali, verticali e obliqui, mentre i demoni cognitivi segnalano la presenza di una specifica lettera. Chiaramente, questo modello ha beneficiato dei risultati neurofisiologici ottenuti dalle ricerche di Hubel e Wiesel, ma non c'è alcun riscontro dell'esistenza di cellule decisionali (la famigerata "cellula della nonna"), inoltre perché i demoni decisionali non siano degli homuncoli sotto mentite spoglie, occorrerebbe precisare l'algoritmo in base al quale funzionano. Ma Selfridge e Neisser al riguardo si limitano a precisare che si tratta di un meccanismo competitivo.
- Il classico concetto neurofisiologico di **cellula della nonna** è così riassunto dalla Churchland (1986): "... l'idea è che livelli successivi implicano l'estrazione di caratteristiche di un tipo sempre più astratto... da cellule che rispondono al contrasto di luminosità a cellule che rispondono ai bordi chiari con una specifica orientazione in una specifica parte del campo visivo,... Più in alto ancora ci sono le cellule che rispondono preferenzialmente a una mano o a una faccia non specifica... Infine, la logica di questo ragionamento ci porta a congetturare l'esistenza di cellule le cui condizioni di risposta siano ancora più astratte e

complesse, come potrebbero essere delle cellule che rispondessero unicamente alla presenza della nonna...”

- La **Teoria del confronto di sagome** (*template matching theory*) assume tre stadi di elaborazione in cascata deputati alla classificazione percettiva (per confronto diretto con sagome presenti in memoria), alla classificazione semantica, e al richiamo del nome. La teoria assume che nella mente ci sia una qualche rappresentazione o sagoma per ciascuna delle configurazioni riconoscibili. La configurazione in entrata viene confrontata con tutte le sagome possibili, per trovare quella che meglio si adatta. Ma, considerando il caso di una “A”, per spiegare il riconoscimento della stessa lettera sottoposta a rotazioni, modificazioni di scala, modificazioni tipografiche, ecc. occorrerebbe aver memorizzato una quantità immensa di sagome, alcune delle quali mai viste prima! I tempi di reazione sarebbero poi enormi. Per questi motivi la teoria è stata presto abbandonata, anche se potrebbe rendere ragione di alcuni fatti osservati come il **repetition priming** (il tempo di reazione nel compito di nominare oggetti visualizzati su uno schermo diminuisce se l’oggetto era stato mostrato precedentemente, ma non se era stato semplicemente denominato).
- La **Teoria dei prototipi** assume che la rappresentazione dell’oggetto in memoria è piuttosto una sintesi statistica di tutte le precedenti esperienze.

Cosa si intende per prototipi? La risposta dipende dal tipo di esperimento. Il prototipo potrebbe essere il pattern dal quale vengono generati vari esemplari distorti da sottoporre al sistema percettivo, oppure il prototipo potrebbe essere una costruzione interna basata sulla percezione di molti pattern di esempio diversi tra loro

Come si formano i prototipi? I meccanismi noti sono

- a) mediazione su tutte le caratteristiche degli esemplari osservati. **Posner, Goldsmith e Weston** (1967) chiedevano ai soggetti di categorizzare liberamente distribuzioni di punti, a loro insaputa ottenuti per deformazione di alcuni patterns prototipo. A una seconda presentazione era anche chiesto di ricordare se ciascuna distribuzione era stata mostrata nella prima prova. Stavolta venivano mostrati anche i prototipi stessi, che correttamente i soggetti assegnavano al gruppo delle loro derivazioni. I soggetti affermavano pure di aver già visto i prototipi. Ciò induceva a ritenere che i soggetti avessero effettivamente “mediato” i pattern nella prima sessione.
- b) selezione delle caratteristiche più frequenti degli esemplari osservati. **Solso e McCarthy** (1981) sottoponevano ai soggetti delle facce che differivano per una o più caratteristiche. Essi osservarono che all’aumentare del numero di caratteristiche per cui differivano le varie facce, diminuiva la capacità di giudicare come “già visti” i prototipi nella seconda sessione.

Quali sono le fasi del processo di riconoscimento del prototipo? Nessuno ha ancora dato una risposta.

- La **teoria del riconoscimento per componenti** di **Irving Biederman** (*Theory of*

Recognition by Components, RBC, 1985,1987,1990) è una importante generalizzazione della teoria del confronto di sagome. Postula l'esistenza di 36 solidi primitivi detti **geoni** (*ioni geometrici*). In realtà con due o tre geoni diversi è possibile rappresentare la maggior parte degli oggetti. Il riconoscimento è il risultato dei seguenti passi

- a) estrazione dei contorni e calcolo dei relativi parametri di forma
- b) definizione dei geoni e delle relazioni geometriche reciproche
- c) confronto con i modelli di oggetti presenti in memoria

Secondo Biederman parti differenti del campo visivo sono elaborate in parallelo. Ciò spiegherebbe perché il tempo di riconoscimento è quasi indipendente dalla complessità della scena.

Biederman ha utilizzato il modello *RBC* per interpretare i rendimenti di soggetti umani in test di riconoscimento visivo. In questi esperimenti la cancellazione parziale dei contorni di una scena determina un crollo delle prestazioni se il contorno cancellato definiva la caratteristica distintiva di un geone.

Nel 1992 Hummel e Biederman hanno implementato la *RBC* su una *NN*.

- L'**approccio ecologico alla percezione** di James J. Gibson (1904-1979) vuole superare la proliferazione dei micromodelli dei processi cognitivi a favore di un modello complessivo del sistema uomo-ambiente; nega che si possa studiare la percezione, o qualsiasi altro processo cognitivo, isolatamente, indipendentemente dal contesto e dal tipo di implementazione, che è la tesi dell'AI e della scienza cognitiva tradizionali; sostiene l'artificialità eccessiva delle indagini psicologiche nel contesto del laboratorio. Di fatto, più si dimostra che il processo percettivo è composto da una serie di microstrutture, più si ha bisogno di un laboratorio per individuarle.

I presupposti filosofici sono quelli del funzionalismo e della teoria dell'evoluzione delle specie.

Secondo Gibson (1966) l'informazione di cui l'organismo ha bisogno ai fini della sopravvivenza è direttamente disponibile in modo non ambiguo nel pattern di luminosità che colpisce l'occhio (**assetto ottico**). Questa informazione non dev'essere intesa nel senso delineato dalla Teoria dell'Informazione di Shannon, né dev'essere estratta dall'input sensoriale per elaborazioni successive, ma sottoforma di **invarianti** (p.es. la forma di un oggetto in avvicinamento) e di **affordances** (la significatività dello stimolo per l'individuo).

Il sistema percettivo sarebbe infatti strutturato per rilevare in modo inconscio e immediato gli *invarianti percettivi* e le *affordances* che sono funzionali all'adattamento della specie in questione.

In "*The Senses Considered as Perceptual Systems*" (1966) Gibson raccoglie le sue idee e descrive i suoi esperimenti sul sistema visivo.

Gli oggetti sono percepiti direttamente (p.es. una sedia) e non secondariamente per composizione di elementi (p.es. pixel, contorni, geoni, ecc.).

Il libro "*The Ecological Approach to Visual Perception*", scritto nel 1979, anno della morte di Gibson, è un testo di riferimento per gli psicologi cognitivisti che

rifiutano o giudicano insufficiente l'approccio dell'AI classica allo studio dei processi cognitivi.

La percezione è istantanea: tra l'ambiente e l'apparato percettivo del soggetto si stabilirebbe una sorta di risonanza, senza necessità di ipotizzare con i cognitivisti dello HIP l'esistenza di rappresentazioni mentali, né con Hebb connessioni progressive delle cellule neuronali.

Il sistema percettivo di una specie è stato modellato dalle forze dell'evoluzione in funzione dell'ambiente, al punto che dovrebbe essere possibile stabilire una corrispondenza biunivoca tra le caratteristiche degli organismi e le caratteristiche dell'ambiente in cui si sviluppano.

Il sistema visivo si è sviluppato per interpretare scene in cui sono presenti sempre molti oggetti contemporaneamente. Gli stimoli semplici non hanno molto senso.

Una scena statica già possiede molti indizi apparentemente secondari ma estremamente informativi: la tessitura del terreno diviene via via più compatta all'aumentare della distanza; il gioco delle ombre rivela i rapporti di distanza tra gli oggetti; più un oggetto è lontano dal punto di osservazione, più si avvicina alla linea di orizzonte, ecc.

Ma come fa l'occhio a ricavare impressioni relativamente statiche di oggetti a partire da proiezioni retiniche in continuo movimento?

Un'**invariante** è una relazione costante tra elementi diversi di una scena statica o nello stesso elemento in fotogrammi diversi di una sequenza

- 1) **invarianti trasformazionali**: p.es. la dimensione apparente di un oggetto in allontanamento si riduce con continuità, e ogni variazione è interpretata come accelerazione dell'oggetto o variazione delle sue dimensioni reali. In questo caso l'invariante è la forma dell'oggetto.
- 2) **Invarianti strutturali**: p.es. il rapporto tra l'altezza di un oggetto e la distanza della sua base dalla linea di orizzonte è un invariante strutturale che ci permette di stimare correttamente la grandezza degli oggetti a distanza.

Le **affordances** mettono in evidenza il valore strumentale degli oggetti percepiti, la loro capacità di offrire una opportunità di interazione motoria con l'ambiente. P.es. un oggetto non è soltanto sferico, ma è "afferrabile". Per Gibson le caratteristiche "afferrabile", "mangiabile", ecc., sono informazioni veicolate dalla luce, proprio come la forma o il colore dell'oggetto.

La teoria ecologica di Gibson è funzionalista, e come tale non si preoccupa di spiegare i meccanismi fisici che guidano, p.es., il calcolo degli invarianti nel cervello.

L'ATTENZIONE

- L'interesse verso i **processi attentivi**, vivo in Donders, Wundt, James, ecc., fu recuperato solo durante la seconda guerra mondiale.

- L'attenzione può essere guidata dagli stimoli, come quando sentiamo un suono particolarmente intenso (processo bottom-up), oppure dalla volontà (processo top-down).
- L'attenzione visiva può essere diretta su una locazione del campo visivo, oppure su uno o più oggetti in movimento (Kanwisher e Driver, 1992).
- Storicamente ha seguito i paradigmi sperimentali dell'*attenzione selettiva* e dell'*attenzione divisa*.
- **Attenzione selettiva** – E' la capacità di selezionare una tra molte sorgenti di informazione contemporaneamente attive (*effetto cocktail party*), mediate dalla stessa modalità sensoriale o da modalità sensoriali diverse. I paradigmi sperimentali utilizzati sono quelli di *selezione* e di *filtraggio*, mentre le metafore adottate sono state:
 - a) attenzione come *filtro*
 - b) attenzione come *serbatoio*
 - c) attenzione come *fascio di luce*
- **Paradigmi di selezione.** In un tipico compito di ricerca visiva al soggetto viene presentato un singolo stimolo (lettera, figura, numero), detto **stimolo target**, seguito dalla presentazione contemporanea di più stimoli, tra i quali il soggetto deve decidere il più rapidamente possibile se è presente o meno lo stimolo target.
- **Paradigmi di filtraggio.** Si basano sulla presentazione continua di stimoli rilevanti (da selezionare) e irrilevanti (da ignorare). P.es. nell'*ascolto dicotico* alla presentazione contemporanea ai due orecchi del soggetto di due diversi messaggi, egli deve prestare attenzione a uno soltanto. Successivamente si potrà chiedere al soggetto di ricordare quanto ascoltato da un orecchio o dall'altro.
- Il problema di stabilire in quale fase dell'elaborazione degli stimoli concorrenti si attui la selezione attentiva distingue *teorie a selezione precoce* e a *selezione tardiva*
- La **teoria del filtro di Donald E. Broadbent** (1958) assumeva una selezione precoce, e si basava sui risultati di Cherry (1953) e di Sperling (1960).
 Al soggetto veniva presentata una sequenza di 6 numeri in modo alternato ai due orecchi. Il soggetto doveva riferire i 6 numeri uditi, nell'ordine desiderato. Quasi sempre il soggetto riferiva prima i numeri uditi da un orecchio, poi gli altri. Anche se istruiti a farlo, i soggetti non erano in grado di ripeterli nell'ordine di presentazione. Secondo Broadbent gli stimoli sono inizialmente analizzati in base a caratteristiche fisiche elementari (p.es. quale orecchio, intensità, ecc.), in parallelo, a partire da un registro presente nel sistema sensoriale (S), in seguito un filtro seleziona lo stimolo da inviare alla memoria a breve termine, attraverso un unico canale percettivo a capacità limitata (P). Il filtro è attivato coscientemente dal soggetto ("voglio prestare attenzione al messaggio che arriva all'orecchio destro") oppure automaticamente (come quando percepiamo un suono particolarmente acuto).
 Una conseguenza della teoria è che negli esperimenti di ascolto dicotico i soggetti possono fornire una prestazione perfetta nel compito di *inseguimento* del

messaggio inviato al canale attentivo, ma possono solo distinguere le caratteristiche fisiche del messaggio inviato all'altro orecchio.

- In effetti, a partire da 1953, nei suoi esperimenti di ascolto dicotico, **Cherry** aveva scoperto che i soggetti potevano distinguere facilmente due voci che differivano per qualche carattere fisico elementare (timbro, intensità, posizione della sorgente), ma non se differivano solo per contenuto semantico. Nella tecnica di **inseguimento** (*shadowing*) Cherry chiedeva al soggetto di ripetere quello che stava ascoltando da un orecchio, mentre un altro messaggio era presentato pure all'altro. Poco o nulla passava del messaggio non inseguito, che era avvertito come un tono puro.
- I risultati di Sperling sulle memorie **ecoica** e **iconica** confermarono l'esistenza del magazzino a brevissimo termine, o registro sensoriale, ipotizzato da Broadbent.
- La **teoria del filtro attenuato**. Ci si accorse tuttavia che in una prova di ascolto dicotico, il soggetto poteva volontariamente cambiare canale attentivo quando chiamato per nome all'orecchio non attentivo (Moray, 1959), oppure spontaneamente quando il messaggio inseguito cambiava canale (Treisman, 1960). Quando Norman (1969) interrompeva il soggetto durante una prova di inseguimento e chiedeva di ripetere quello che aveva udito dal canale non attentivo, il soggetto era in grado di ripetere l'ultima parola. Ciò portò a modificare la teoria di Broadbent ipotizzando che il filtro si limitasse ad attenuare e non a bloccare l'informazione disponibile nei canali non attentivi. Secondo la teoria del filtro attenuato, proposta da **Anne M. Treisman** (1960,1969) e adottata nel 1971 anche da Broadbent, i canali percettivi sono molti (P1,P2,..) e differenziati per caratteristiche fisiche, lessicali e semantiche. Ogni canale ha associata una **enfasi** variabile nel tempo e una soglia. Ciascun canale trasferisce l'informazione relativa alla memoria solo se la sua attivazione supera questa soglia.
- Le **teorie della selezione tardiva**, come quelle di Deutsch e Deutsch (1963), Norman (1968), Moray (1975) e La Berge (1975), affermano che tutti i canali indistintamente arrivano ad essere analizzati fino ai livelli percettivi e discriminativi. Solamente l'accesso alla coscienza, alla memoria a lungo termine e ai sistemi di risposta comportamentale, sarebbe regolato in modo differenziato per i diversi canali, in ragione di appositi coefficienti. Ciò sulla base di due indizi
 - 1) Esistenza di una **elaborazione semantica di stimoli non attesi**. Tuttavia l'effetto è troppo piccolo (solo il 16% degli stimoli inattesi è elaborato anche semanticamente)
 - 2) L'**effetto categoria**. Si riferisce alla maggiore facilità con cui si identifica una lettera in un insieme di numeri, rispetto al caso che sia mescolata con altre lettere. In tal caso i numeri verrebbero percepiti come una sorta di sfondo omogeneo, inducendo a ritenere che un'analisi semantica in parallelo si verifichi per ciascuno di essi.

- L'**attenzione divisa** fa riferimento alla capacità di prestare attenzione a più cose contemporaneamente.
- Secondo la metafora dell'attenzione come *sebatoio*, l'individuo avrebbe a disposizione una quantità limitata di **risorse attentive**, o di capacità di elaborazione dell'informazione.
- Per quanto riguarda l'utilizzo di tali risorse, alcuni ritengono che queste debbano spostarsi in toto da un compito all'altro (**teorie strutturali**), altri pensano che possano essere distribuite (**teorie della capacità**).
- Il **modello di Kahneman** (1973) tenta di unificare le teorie strutturali e quelle di capacità.
 - 1) l'attenzione è limitata, anche se il limite è variabile
 - 2) fino al suo limite l'attenzione cresce linearmente con le richieste del o dei compiti
 - 3) l'attenzione è divisibile, ma a livelli elevati di carico tende a passare in modo unitario da un compito all'altro
- **Norman e Bobrow** (1975) ipotizzano la **funzione prestazione-risorse (PRF)** che mette in relazione la prestazione con le risorse disponibili. Essi distinsero anche **compiti limitati dalle risorse e compiti limitati dai dati**. Un esempio di compito limitato dai dati è la comprensione di un discorso in una lingua straniera a noi poco familiare: per quante risorse investiamo la prestazione sarà scadente, per mancanza di dati percettivi o mnestici.
- La **teoria delle risorse multiple** (Navon e Gopher, 1979; Sanders, 1979; Wickens, 1984) introduce un ulteriore distinguo nel *tipo di risorse*. Ciò per spiegare i seguenti fenomeni:
 - a) l'insensibilità alla difficoltà: l'incremento di difficoltà di un compito primario non riduce la prestazione del compito secondario
 - b) la condivisione perfetta: due compiti vengono eseguiti contemporaneamente senza alcuna interferenza, anche se ciascuno singolarmente può interferire con un terzo
 - c) l'alterazione strutturale: cambiando la struttura del compito (p.es. la modalità di risposta), ma non la sua complessità, può determinare aumento d'interferenza col compito secondario

Dagli esperimenti di Wickens (1963) era noto che più gli elementi contenuti nella MBT sono simili, più interferiscono tra loro, peggiorando le prestazioni di richiamo.

Secondo Wickens le risorse mentali possono essere caratterizzate da tre dimensioni differenti:

- 1) modalità di input e di output
- 2) tipo di elaborazione (p.es. verbale, spaziale)
- 3) stadio di elaborazione in cui è richiesta la risposta

Secondo il modello di Wickens, più diverse sono le risorse richieste per eseguire due compiti, minore sarà l'interferenza, e più efficiente sarà la divisione del lavoro.

P.es. il compito di guidare (input visivo, elaborazione spaziale, output motorio) interferisce poco con quello di colloquiare con il passeggero accanto (input uditivo, elaborazione verbale, output verbale).

Il fatto è che spesso nella vita capita che i compiti interferenti coinvolgano la stessa modalità sensoriale. In questi casi lo studio dell'attenzione sulla base della metafora della luce pare più fruttuoso.

- Nello studio di **Yarbus** del 1967 sui movimenti oculari durante la percezione visiva di una scena complessa, risultava che l'ispezione oculare era un'alternanza di fissazioni e di saccadi. Ma come fa il sistema visivo a decidere quale sarà il prossimo punto da fissare? Si è ipotizzato che esistesse un **processo attentivo**, di natura seriale e focale, guidato da un **processo pre-attentivo** non focale, che evidenzia le zone del campo visivo più interessanti.
- **Eriksen e Yeh** (1985) hanno proposto per primi la **metafora del fascio di luce** (*spotlight*) secondo cui nel processo pre-attentivo l'attenzione sarebbe concentrata solamente nell'intorno di un punto, il **focus attentivo**. Questo si comporta come un fascio di luce: si sposta in modo continuo (Shulman *et al.*, 1979), a velocità costante (Tsal, 1983), ha un diametro variabile, la velocità di elaborazione per stimoli presentati al suo interno è maggiore che per stimoli presentati in altre locazioni spaziali. Sembra inoltre che l'attenzione non può facilmente essere allocata in regioni non adiacenti del campo visivo (Eriksen e Yeh, 1985). Secondo Eriksen e Yeh il posizionamento del focus attentivo sarebbe guidato dagli stessi stimoli visivi (*bottom-up*), ma anche da livelli cognitivi più elevati (*top-down*).
- Gli studi di **Posner** sull'attenzione selettiva visiva (1978, 1988, 1990) confermarono e perfezionarono la metafora. Il focus attentivo non coincide necessariamente col **punto di vista**, il quale negli esperimenti di Posner si trova sempre al centro dello schermo. Nel **paradigma di Posner** (1978) ogni prova consiste nella presentazione di uno stimolo segnale (*cue*), ">", "<", "<>", al centro di uno schermo diviso in due metà. Lo stimolo segnale predice in quale zona dello schermo verrà successivamente presentato uno stimolo target, che il soggetto dovrà identificare più rapidamente possibile premendo un pulsante. Lo stimolo segnale, quando non neutro ("<>"), può essere vero (*prova valida*, 80% dei casi) oppure falso (*prova invalida*, 20% dei casi). Normalmente si osserva, rispetto alla condizione neutra, una riduzione dei tempi di risposta nelle prove valide e un aumento nelle prove invalide. Dato che il soggetto doveva sempre mantenere fisso il punto di vista al centro dello schermo, i risultati di Posner confermano che il sistema pre-attentivo è guidato in una certa zona del campo visivo dalla conoscenza dello stimolo segnale, con un processo *top-down*.
- Le ricerche di **Navon** (1977) precizarono che l'apertura dell'intorno attentivo è dapprima grande, poi si riduce. Contrariamente a ciò Eriksen e Yeh pensavano che tale apertura fosse regolabile dalle aspettative del soggetto con un processo *top-*

down, e che tale apertura fosse inversamente proporzionale alla risoluzione visiva (una sorta di zoom). Nei suoi esperimenti Navon mostrava ai soggetti una lettera grande (p.es. A) formata da lettere più piccole, dello stesso tipo (p.es. A, *stimoli congruenti*) o di tipo diverso (p.es. E, *stimoli non congruenti*). Mentre i tempi di reazione nel riconoscimento delle lettere grandi era lo stesso per stimoli congruenti e incongruenti, le prestazioni nel riconoscimento delle lettere piccole erano migliori con gli stimoli congruenti. Inoltre il riconoscimento era sempre più efficiente per le lettere grandi.

- In realtà la questione è ancora controversa, infatti pare che a volte il sistema pre-attentivo sia guidato dalle aspettative, altre volte dallo stimolo.
- Nel ***paradigma della ricerca visiva*** il soggetto deve trovare uno o più stimoli target all'interno di un certo pattern di stimolazione, dovendo mantenere lo sguardo fisso su un punto, oppure potendo eseguire movimenti oculari. Distinguiamo il ***paradigma dei tempi di reazione*** (lo stimolo target è solo uno, si deve decidere se lo stimolo target è presente o meno il più rapidamente possibile, il tempo di esposizione del pattern è lungo a piacere) dal ***paradigma dell'accuratezza*** (gli stimoli target possono essere molti, dopo un tempo detto SOA si deve dire quanti stimoli target erano presenti nel pattern, il tempo di esposizione del pattern è tanto breve da impedire movimenti oculari).
- In entrambe i paradigmi si osserva che le prestazioni sono migliori quando lo stimolo target è facilmente distinguibile dagli stimoli distrattori
- La Treisman (1977) chiedeva ai soggetti di localizzare uno stimolo target tra numerosi stimoli distrattori. Scopì che quando lo stimolo target differiva per una sola caratteristica elementare dagli stimoli distrattori (forma, colore, orientamento) il tempo di individuazione era indipendente dal numero di stimoli distrattori. Se invece vi erano distrattori maggiormente diversificati il tempo di risposta aumentava in proporzione al loro numero. Secondo il modello proposto dalla Treisman, gli stimoli sono processati in parallelo su tutto il campo visivo se il target è definito da un'unica caratteristica distintiva, altrimenti la ricerca visiva è di tipo seriale (un item alla volta).
- Su questo modello la Treisman ha fondato poi la ***Feature Integration Theory*** (1980), secondo cui il sistema attentivo sarebbe formato da un sistema pre-attentivo di rilevatori, uno per caratteristica distintiva del target. Ciascun rilevatore opera in parallelo su tutto il campo visivo e determina una mappa delle locazioni degli stimoli che hanno quella caratteristica. Se le caratteristiche del target sono molte, altrettante mappe dovranno integrarsi tra loro in modo che un processo di ricerca seriale possa identificare lo stimolo che le possiede tutte.
- In definitiva si pensa che esistano due modi di funzionamento dell'attenzione visiva: uno focale e seriale che seleziona soprattutto la locazione spaziale, l'altro parallelo e distribuito che seleziona soprattutto una determinata caratteristica.
- Quali caratteristiche sono elementari?
Def. Una caratteristica è elementare se

- a) la ricerca visiva dell'item dà tempi di reazioni indipendenti dal numero degli item
- b) regioni del pattern contenenti solo item di questo tipo sono segregati già nel processo pre-attentivo

esempi: colore, orientamento, stato di moto. Altre caratteristiche come curvatura, parallelismo, dimensione, freq.spaziale, forma, non si sa.

- Esistono performance intermedie rispetto alle due distribuzioni?

E' ampiamente dimostrato che esiste un continuum di performance intermedie (p.es. Zohary, Hochstein 1989).

Ad alimentare il discredito sulla dicotomia tra processi seriali e paralleli, hanno contribuito alcuni esperimenti che mostrano come i risultati di performance che sembrano dovuti a una ricerca solo seriale possono essere previsti da un modello di ricerca solo parallelo con risorse a capacità limitata. P.es. Ratcliff (1978) le risorse attentive R vengono distribuite equamente su tutti gli N items. Ciò avverrebbe cumulativamente ad ogni scansione parallela del campo visivo, fin quando almeno un item non raggiunge un grado di evidenza E maggiore di una soglia S . Sicché

$$E_i = E_{i-1} + R/N > S$$

$$E_0 = 0$$

da cui, nel continuo,

$$E(t) = (R/N)t > S$$

Il tempo di reazione sarà quindi proporzionale a

$$t = SN/R$$

cioè proporzionale al numero degli item.

Palmer e McLean (1995) hanno anche modificato il modello in modo da eliminare l'ipotesi delle risorse limitate. In definitiva sembra superfluo postulare l'esistenza di un processo di ricerca seriale. Oggi si preferisce parlare di ricerca "efficiente" e "poco efficiente" piuttosto che di ricerca parallela e seriale.

- Che ruolo hanno i distrattori?

Duncan e Humphreys (1989) hanno scoperto che l'efficienza della ricerca visiva diminuisce all'aumentare della somiglianza tra item target e distrattori, e aumenta all'aumentare della somiglianza tra item distrattori. Resta non è ben definita la nozione di somiglianza.

- Come fanno i soggetti a distinguere item diversi?

Questo problema è connesso alla dicotomia tra analisi globale e analisi locale. Secondo **Nakayama** (1990) la scala di analisi dipende dal compito. Se M bit è la massima informazione ottenibile su una fissazione attentiva e C bit sono necessari

a definire le caratteristiche di un item, con una fissazione si possono distinguere $K=M/C$ items. Se C aumenta (compito più difficile), ma il numero totale di items resta uguale nel pattern, occorre restringere il campo attentivo su regioni più piccole del pattern. Nel 1992 Verghese e Pelli hanno stimato $M=50$ bit.

- I processi pre-attentivi sono influenzati in modo top-down?
E' dimostrato che esistono effetti di apprendimenti precedenti sulle performance nei compiti di ricerca visiva.
- L'impressione che abbiamo di vedere continuamente e completamente anche nei dettagli una scena è illusoria: in parte grazie alla presenza dei continui movimenti oculari, in parte per l'uso che facciamo della memoria, dato che normalmente poco o nulla della scena cambia in un breve lasso di tempo.

PENSIERO E SOLUZIONE DEI PROBLEMI

- Il pensiero è la capacità di astrazione, di rappresentazione della realtà anche in assenza di essa, di pianificazione del comportamento, di estrazione delle regolarità nei fenomeni, di previsione, di inferenza e deduzione.
- Se il pensiero è cosciente il soggetto è in grado di definire tutti i passaggi del suo sviluppo
- Nell'ambito della Psicologia Cognitiva i processi di pensiero sono costrutti ipotetici usati per spiegare certi tipi di comportamenti, come la capacità di risolvere problemi.
- Nel 1920 **Max Wertheimer** introdusse la distinzione tra ***pensiero riproduttivo***, in cui si tenta di risolvere un problema applicando meccanicamente delle regole, e ***pensiero produttivo***, in cui si giunge a una soluzione intuitiva per ristrutturazione della situazione problematica. Molti esempi da lui raccolti furono pubblicati postumi in *Productive Thinking* (1945).
- Il ***pensiero produttivo*** accresce la conoscenza senza incremento dei dati a disposizione, in quanto è la modifica dello schema di relazioni tra gli stessi dati. Dal punto di vista comportamentale si esprime attraverso l'uso di strategie di soluzione completamente diverse da quelle usate fino a un momento prima.
- Oggi si accetta che una risposta comportamentale alla soluzione di un problema sia prima emessa "internamente", a seguito di una serie di simulazioni mentali di ricerca. Questi processi non sono osservabili, se non dal soggetto medesimo.
- I problemi da risolvere nei diversi paradigmi sperimentali sono
 - a) di organizzazione dei mezzi a disposizione (p.es. gli anagrammi)
 - b) di scoperta della struttura
 - c) di trasformazione

Problemi di organizzazione

- Tra il 1913 e il 1917 **Wolfgang Kohler** effettuò degli studi sull'intelligenza degli scimpanzé come direttore dell'*Accademia prussiana delle scienze* dell'isola di Tenerife, nell'arcipelago delle Canarie. Egli osservò come gli scimpanzé, posti di fronte a un problema, come afferrare una banana posta al di fuori della gabbia, riuscivano a trovare una soluzione intelligente (p.es. aiutandosi con un bastone e una cassa) non per approssimazioni successive per prove ed errori, ma quasi con un atto di illuminazione improvviso, che egli chiamò **Insight** (*einsicht*, “vedere dentro”).
- Le successive precisazioni di **Karl Dunker** (1935) rappresentano un tentativo di rispondere a un problema: perché la percezione visiva di una situazione-problema non porta automaticamente alla soluzione? Egli osservò che:
 - a) il campo problemico non è sempre di natura ottica, talvolta è di tipo logico non facilmente visualizzabile
 - b) spesso nei problemi logici intervengono forze che impediscono di percepire alcuni oggetti in relazioni diverse da quelle che siamo abituati a considerare (**fissità funzionale**)
 - c) L'abitudine ci porta ad applicare a un problema nuovo un metodo che si è rivelato utile in passato su un problema apparentemente simile
 - d) spesso, come avviene nella ricerca scientifica, nel campo problemico si possono avere solo **insight parziali** (problemi complessi possono richiedere trasformazioni successive).
- L'assunto di base del **comportamentismo** è che i processi di soluzione dei problemi siano basati sull'uso di associazioni **stimolo-risposta**, derivanti da rinforzi ottenuti in passato. La strategia di fondo è quindi quella per **prove ed errori**.
- **Hull** ha raffinato questo punto di vista, distinguendo meccanismi **divergenti** (uno stimolo è associato con forze diverse a più risposte) e **convergenti** (più stimoli sono associati con forze diverse alla stessa risposta). Hull chiama **abitudine** l'associazione S-R. Una **famiglia di abitudini** è una sequenza comportamentale complessa formata da più connessioni S-R, avente una sua forza di attivazione che la colloca su un livello specifico di una gerarchia.

Problemi di scoperta della struttura

- Esempi sono: determinare la cifra da aggiungere alla successione numerica ..., oppure completare la seguente frase ...
- I problemi di analogia, $X:Y=Z:?$ (p.es. cane abbaia come gatto ?) richiedono il ragionamento analogico, poco studiato in Psicologia Cognitiva. Uno dei modelli più noti è quello di Stenberg (1977), che distinse tre sottoprocessi:
 - 1) Codifica. Definizione degli attributi rilevanti
 - 2) Inferenza. Definizione delle relazioni valide tra X e Y
 - 3) Corrispondenza. Definizione delle relazioni valide tra X e Z
 - 4) Applicazione. Tentativo di individuazione di ? tra possibili candidati {A,B,..}.

Problemi di trasformazione

- Partendo da uno stato iniziale si deve passare a uno stato finale tramite una serie di operazioni (p.es. scacchi, dama, torre di hanoi).

RAPPRESENTAZIONE DELLA CONOSCENZA IN PSICOLOGIA COGNITIVA

Mappe cognitive

- Le **mappe cognitive** sono state introdotte da **Tolman** (1932), di cui si è già discusso in precedenza. Dai numerosi studi svolti su soggetti umani sembra che le mappe cognitive siano una sorta di mappa geografica, con punti salienti (**nodi**) connessi da relazioni di distanza, non fedeli al reale. Non è chiaro se le mappe cognitive siano rappresentate come sequenze di *percorsi* (**route**) o globalmente come una vista dall'alto (**survey**), né se le coordinate siano riferite al soggetto (**egocentriche**) o no (**allocentriche**).
- Le **immagini mentali** sono una rappresentazione interna della struttura spazio-temporale degli eventi. I **proposizionalisti** (come **Zenon W. O'Lyshyn**) sostengono che si tratti di descrizioni solo proposizionali, gli **immaginatisti** (come **Stephen Michael Kosslyn**) ammettono anche l'esistenza di descrizioni pittoriche, vere e proprie immagini. L'isomorfismo tra immagini mentali e oggetti reali ammette comunque deformazioni. Nella seconda metà degli anni '60 l'immaginatista **Roger Shepard** ha eseguito una serie di esperimenti volti a valutare
 - 1) effetti di *distanza* (isomorfismi sulla distanza euclidea tra stimoli visivi)
 - 2) effetti di *grandezza* (tempo di detezione di un oggetto)
 - 3) effetti di *trasformazione* (rotazioni, traslazioni)ùLa variabile dipendente è sempre il tempo impiegato dal soggetto a svolgere una determinata operazione mentale (**cronometria mentale**)
Noto è l'esperimento in cui il soggetto deve decidere se una data figura è ottenuta per rotazione di una figura di base. Molti autori hanno trovato che il tempo di reazione è direttamente proporzionale all'angolo di rotazione.

I concetti

- I **concetti** sono una delle rappresentazioni della conoscenza di più alto livello. Sul problema della strutturazione dei concetti distinguiamo le teorie:
 - a) *classiche*
 - b) *dei tratti differenziali*
 - c) *dei prototipi*

Le **teorie classiche** risalgono ad Aristotele. Un concetto l'insieme degli oggetti che rispettano le sue caratteristiche definitorie (esemplari), che sono a loro volta concetti. Quindi i concetti si uniscono a formare nuovi concetti, per mezzo di connettivi logici come “e”, “oppure”, “se-allora”, “se-e-solo-se”, ecc., secondo un meccanismo automatico indipendente dall'esperienza. La teoria classica assume che le difficoltà di apprendere un concetto è proporzionale solo al numero delle caratteristiche definitorie. Questo fatto è smentito da recenti **esperimenti di apprendimento dei concetti**. P.es. nel paradigma di **ricezione** si sottopone al soggetto una successione di esemplari (tipicamente figure) finché questi non indovina il concetto che li unisce. Si è visto che i connettivi logici tra due caratteristiche A e B non hanno tutti uguale facilità di essere compresi. Il più difficile è “A se-e-solo-se B”, il più facile “A e B”. Queste differenze tendono però ad attenuarsi con la pratica.

Le **teorie dei tratti differenziali** asseriscono che l'attribuzione di un dato esemplare a un dato concetto avviene per somiglianza. Ogni concetto ha, oltre alla lista delle **caratteristiche definitorie**, la lista delle **caratteristiche tipiche**, cioè più frequenti. La valutazione di somiglianza è fatta prima sulle caratteristiche tipiche. Se questo primo confronto non porta a una decisione si passa a considerare in modo sequenziale le caratteristiche definitorie. La teoria prevede una maggiore velocità nel riconoscere esemplari tipici, risultato confermato negli studi sulla memoria semantica.

- Quando vogliamo dare una definizione di un oggetto qualsiasi, iniziamo esprimendo la categoria di appartenenza. **Collins e Quillian** (1969) proposero un modello di **memoria semantica** strutturato ad albero, in cui ogni nodo rappresenta una categoria (di oggetti, animali, ecc.) che appartiene alla categoria rappresentata dal nodo padre. Inoltre, a ogni nodo sono associate delle proprietà particolari. Per verificare la realtà psicologica del modello gli autori sottoponevano a volontari frasi come “*uno squalo è un pesce*”, “*uno struzzo è un pesce*”, ecc. alle quali esso dovevano rispondere *vero* o *falso*. Si osservò che effettivamente la latenza misurata era proporzionale alla distanza tra le categorie nell'albero semantico (**category size effect**).
- Successivamente **Rips, Shoben e Smith** (1973) misero in luce che la distanza semantica soggettiva è in generale diversa da quella “ufficiale”, sicché cane e mucca erano giudicati più vicini al concetto di animale che a quello di mammifero. Inoltre la distanza semantica tra *uccello* e *usignolo* era giudicata minore rispetto alla distanza *uccello* e *pinguino* (**effetto tipicità**). Secondo Rips l'**effetto tipicità** si spiega assumendo che la ricerca nella memoria semantica si basi inizialmente non sui quegli attributi comuni a tutti gli elementi della classe (**proprietà definitorie**) ma sugli attributi più frequenti (**caratteristiche**). Da questo punto di vista l'*usignolo* ha molte caratteristiche dell'*uccello* tipico (vola, emette suoni modulati, ha una mole media), mentre il *pinguino* no.
- L'oggetto tipico, che contempla tutte le **caratteristiche** è detto **prototipo**.

- Le **teorie dei prototipi** prendono le mosse dall'obiezione che per alcuni concetti è quasi impossibile dare una lista di caratteristiche, p.es., come già notò Wittgenstein, per il concetto di "gioco". Negli anni '80 molti ricercatori hanno così adottato un punto di vista completamente opposto, secondo cui non esistono rappresentazioni globali di categorie, ma solo raggruppamenti di esemplari concreti. Alcuni autori assumono che i prototipi sarebbero delle tendenze centrali delle categorie, che però non coincidono con nessuno degli esemplari della categoria (**teorie dei prototipi-prototipi**). Altri assumono che i prototipi sono esemplari anch'essi (**teorie dei prototipi-esemplari**).
- Secondo la teoria dei prototipi-prototipi della studiosa americana **E. Rosch** (1973) tra tutti i concetti quelli che hanno un prototipo ben definito (l'esemplare più rappresentativo) sono chiamati **categorie** (p.es. "frutto", "uccello", "auto", ecc.), per gli altri è difficile pensare a un prototipo (p.es. il concetto di "libertà"). Le categorie non sono definite da caratteri dicotomici, ma hanno una struttura formata da un "**nocciolo**" (il *prototipo*), circondato da esemplari diversi, a distanza dal prototipo inversamente proporzionale alla somiglianza con questo. Gli esemplari più distanti possono essere inclusi nei **cerchi di prototipicità** di altre categorie. Inoltre non esistono *proprietà definitorie* uguali per tutti, ma proprietà più frequenti tra i vari esemplari di una categoria. Si distinguono categorie sovraordinate (p.es. uccello, mobile), di base (p.es. canarino, sedia) e subordinate (p.es. il mio canarino Titì, sedia a dondolo). Le categorie di base sono quelle più generali cui è facilmente associabile un'immagine, sono le più usate nel linguaggio comune e le prime che il bambino utilizza. I paradigmi sperimentali usati dalla Rosch sono stati
 - a) **valutazioni di tipicità**. Il soggetto deve valutare quanto un dato esemplare è tipico della categoria sovraordinata, o deciderne l'appartenenza più rapidamente possibile
 - b) **valutazioni di similarità**. Il soggetto deve valutare il grado di similarità di due esemplari di una stessa categoria sovraordinata

Si è osservato che in ogni categoria ci sono esemplari più tipici, per i quali il tempo di decisione di appartenenza è minore e che sono nominati per primi in un compito di elencazione libera. Per gli esemplari di una stessa categoria le valutazioni di tipicità sono correlate positivamente con la somiglianza di famiglia (concetto introdotto da Wittgenstein secondo cui non si decide l'appartenenza di un esemplare a una categoria per confronto delle caratteristiche, ma in base alla somiglianza con altri esemplari già considerati come facenti parte della categoria in questione).

Def. Se una categoria C è un insieme $\{e_i\}$ di esemplari e_i nello spazio delle caratteristiche $\{0, 1\}^M$, La **somiglianza di famiglia** di un esemplare e in C è

$$\sum_{k=1..M} e_k (\sum_i e_{ik})$$

James A. Hampton (1993) definisce il **peso prototipico** associato alla caratteristica k -esima come la sua presenza media in C

$$w_k = (1/|C|) \sum_i e_{ik}$$

Per decidere se un prototipo s appartiene a C Hampton ha introdotto la **somiglianza**

$$S(C, s) = \sum_{k=1..M} w_k s_k$$

Una volta calcolate tutte le somiglianze $S(C_p, s)$, si decide per la categoria C che massimizza la probabilità (regola di **Luce**)

$$P(C_p, s) = S(C_p, s) / [S(C_p, s) + \sum_{q \neq p} S(C_q, s)]$$

Chiaramente, una volta deciso che $s \in C$, tutti i pesi prototipici devono essere riaggiornati.

- La teoria dei prototipi-esemplari cercano di superare i limiti che le teorie dei prototipi-prototipi hanno con le categorie non linearmente separabili, per le quali i pesi prototipici sono tutti uguali. P.es. nello spazio definito dalle caratteristiche {triangolo, quadrato, verde, rosso}, la categoria “quadrato se e solo se rosso” contiene gli esemplari $C = \{\text{quadrato rosso, triangolo verde}\}$. I pesi prototipici delle caratteristiche in C sono tutti uguali a 0.5. Questo significa che qualunque esemplare ha somiglianza 1 con C , e non è mai possibile deciderne l'appartenenza a C .

La **teoria del contesto** ideata da **Medin e Schaffer** (1978) prevede che la somiglianza sia definita come

$$S(C, s) = \sum_i S(e_i, s)$$

cioè in termini delle somiglianze con gli esemplari già in C

$$S(e_i, s) = \prod_{j=1..M} \text{if}(e_{ij} = s_j, 1, m)$$

I parametri $0 < m_j < 1$ sono detti di **attenzione selettiva**. Se m_j è prossimo a 0 la j -esima caratteristica ha grande importanza ai fini della decisione di appartenenza di s a C . Nosofsky ha affinato il modello supponendo che m cambi con l'esperienza. Riprendendo l'esempio fatto in precedenza, si ottiene che la somiglianza di un esemplare non in C vale m^2 , mentre la somiglianza con uno in C vale $m^4 + 1$. Per $m \neq 1$ si ha $m^4 + 1 > m^2$, e la scelta di una soglia opportuna ci fa decidere correttamente per l'appartenenza.

Gli schemi

- Fu Barlett nel 1932 a introdurre vagamente il concetto di schema per evidenziare l'organizzazione della memoria nella comprensione e nel ricordo di storie. Nel 1977 Rumelhart e Ortony hanno definito lo schema come una struttura cognitiva con
 - 1) Insieme di variabili di cui occorre specificare i valori
 - 2) Possibilità di introdurre altri schemi, gerarchicamente
 - 3) Diversi livelli di astrazione delle informazioni contenutevi

Appunti integrati su
RETI NEURALI E PROCESSI COGNITIVI
 E.Pessa
MODELLIZZARE LE FUNZIONI DEL CERVELLO
 Daniel J.Amit

CENNI STORICI

- **W.S.McCulloch e W.Pitts** (1943), per spiegare il funzionamento logico della corteccia cerebrale, proposero un modello *esplicito* di rete di **neuroni formali** che a loro avviso
 - a) poteva realizzare qualsiasi funzione booleana
 - b) è equivalente ai *FA* (automi finiti)
- Negli anni '30 **N.Rashevsky** aveva ideato una descrizione *implicita* del funzionamento dei neuroni, basato su equazioni differenziali
- Fin dai primi anni si intuisce che l'apprendimento può essere associato a una variazione di *W* per azione di segnali provenienti dall'esterno o dell'attività interna della rete.
- Nel 1949 D.Hebb avanza l'ipotesi della **facilitazione sinaptica** (o **legge di Hebb**): ogniquale volta l'eccitazione di un neurone A è immediatamente seguita (dopo un tempo pari al ritardo sinaptico) dall'eccitazione di un neurone B, aumentava l'efficacia sinaptica della connessione tra l'uscita di A (l'assone) e l'ingresso di B (dendriti).
- Nel 1952 la descrizione implicita del neurone riceve forte impulso dalla scoperta delle equazioni di propagazione del potenziale d'azione da parte di **A.L.Hodgkin e A.F.Huxley** (HH).
- Nel 1962 **F.Rosenblatt** costruisce il **Perceptron**, estensione naturale del neurone formale di McCulloch e Pitts. Sostanzialmente viene aumentato il numero di ingressi, da uno o due a un numero arbitrario. Era capace di apprendere sulla base del confronto tra uscita e risposta attesa fornita da un supervisore (lo stesso sperimentatore).

- **B.Widrow e M.Hoff** ottimizzano l'algoritmo di apprendimento supervisionato su una rete a due strati e connessioni feedforward. Le unità hanno funzione di attivazione arbitraria.
- Nel 1961 **E.R.Caianiello** dà una formulazione generale della dinamica di reti con unità a soglia e ritardi sinaptici qualunque e mostra come possano nascere attività riverberanti
- Nel 1961-2 **R.FitzHugh, J.Nagumo, S.Arimoto e S.Yoshizawa** propongono una descrizione semplificata del modello di HH, ma che ne conserva le proprietà salienti.
- Negli anni '60 i successi dell'AI pongono in ombra la ricerca sulle NN. Il libro "*Perceptrons*" di **M.Minsky e S.Papert** (1969) dimostra che esistono problemi di classificazione che il Perceptron non può risolvere. Segue un periodo di tagli finanziari e oblio per le NN.
- Ma dal 1970 **P.A.Annios** sviluppa la neurodinamica statistica, studio della dinamica della distribuzione di probabilità di attivazione di una classe di reti.
- Nel 1972 **A.Amari** connette la nozione probabilistica di campo di attività della rete con il funzionamento delle singole unità, la prima sintesi tra descrizione esplicita e implicita.
- Nel 1973 **Von der Malsburg** inventa un modello di rete sede di processo di auto-organizzazione, dove i coefficienti sinaptici variano in funzione della correlazione temporale tra le attività dei nodi. Il suo intento era spiegare la formazione delle ipercolonne scoperte nella corteccia visiva
- Nel 1973 **H.R.Wilson e J.D.Cowan** introducono un sistema di equazioni differenziali per spiegare l'attività del tessuto nervoso corticale e talamico.
- Dal 1969 **S.Grossberg** propone un modello multistrato auto-organizzanti, con connessioni ricorrenti. Rileva importanti analogie con processi di percezione, apprendimento, pensiero, controllo motorio, ecc.
- **J.A.Andersen** espone il modello **BSB** (*Brain State in a Box*), con neuroni dotati di saturazione e connessioni di feedback tra unità di uscita e di ingresso.
- Nel 1981 **T.Kohonen** introduce il modello più semplice di rete auto-organizzanti.
- **W.A.Little e G.L.Shaw** dimostrano dal punto di vista statistico l'equivalenza formale tra una rete di unità binarie a soglia e i **vetri di spin**, sistemi magnetici che si formano, ad esempio, introducendo in un metallo impurità di un altro metallo. Storicamente i primi a individuare l'analogia furono **Cragg e Temperley** (1954). Tipicamente un metallo ha una temperatura critica, al di sopra (sotto) della quale è in uno stato paramagnetico (ferromagnetico). Un vetro di spin invece a bassa temperatura ha stati di equilibrio stabili dove i singoli atomi possono avere momenti magnetici con direzioni e versi disordinati.
- Nel 1982 **J.Hopfield** dimostra come questi sistemi possano fungere da memorie associative, con leggi di attivazione e di apprendimento molto semplici, e inaugura il paradigma delle macchine connessionistiche. In questo periodo l'AI colleziona i primi insuccessi.

- In ambito psicologico, *J.A.Feldman e D.H.Ballard* (1982) propongono modelli di processi cognitivi basati sui promettenti sviluppi delle NN, i **modelli connessionistici**.
- Nel 1985 *D.E.Rumelhart, G.E.Hinton e R.J.Williams* introducono una nuova regola di apprendimento supervisionato, detta **error back-propagation**, capace di superare i limiti del Perceptron, e di dotare le reti di capacità di apprendimento teoricamente illimitate.
- Del 1986 è l'antologia in due volumi di J.D.McClelland e D.E.Rumelhart, *Parallel Distributed Processing*, che getta le basi per un approccio subsimbolico, parallelo e distribuito all'AI.
- Una suddivisione dei vari modelli di NN è quella tra **reti feed-forward** (esiste un legame funzionale tra ingresso e uscita, ingressi diversi producono uscite diverse) e **Attractor NN** (l'output è l'attrattore raggiunto, ingressi diversi possono convergere sullo stesso attrattore). Nel primo gruppo gli associatori lineari, il perceptrone multiplo, nel secondo gruppo le reti booleane, di Caianiello (1961), Amari (1972), Grossberg (1969), Little (1974), Hopfield (1982).
- Le NN rappresentano la realizzazione più generale del concetto di sistema, nel senso che qualunque sistema è rappresentabile sotto forma di NN. Consideriamo NN a tempo discreto.
- Fra i tipi di **unità** maggiormente usati:
 - 1) unità **binarie a soglia**. L'uscita è pari a 1 se il **potenziale di attivazione** è maggiore o uguale a una **soglia**, altrimenti è 0,

$$P(t) = \sum_{i=1..N} w_i x_i(t)$$

$$u(t+1) = \Psi(P(t)-S)$$

dove Ψ è la funzione gradino.

Def. Sono reti binarie tutte quelle con uscita delle unità a 2 valori. Le unità con output $u=\pm 1$ sono particolarmente attraenti per i fisici, in quanto direttamente collegabili al modello di Ising dei sistemi magnetici. Le rappresentazioni $u \in \{0,1\}$ e $v \in \{-1,1\}$ sono equivalenti, essendo legate dalla trasformazione $u = \frac{1}{2}(v+1)$.

$$v(t+1) = \text{sgn}(P(t)-S), \quad P(t) \neq S$$

- 2) unità booleane. L'uscita è una funzione booleana degli n ingressi
- 3) unità **lineari**. Con soglia

$$u(t+1) = k(P(t) - S) \Psi(P(t) - S)$$

o senza soglia

$$u(t+1) = kP(t)$$

4) unità **con uscita sigmoideale**.

$$u(t+1) = 1/[1 + e^{-(P(t)-S)}]$$

oppure

$$u(t+1) = 1/2 + \arctg[P(t)-S]/\pi$$

Talvolta l'uscita assume valori $\{-1, 1\}$,

$$u(t+1) = -1 + 2/[1 + e^{-(P(t)-S)}]$$

oppure

$$u(t+1) = \tanh[P(t)-S]$$

5) unità **$\Sigma\Pi$ (sigma pi)**. Il potenziale è una somma di prodotti degli ingressi. P.es.

$$P(t) = p_1 x_1(t)x_2(t) + p_2 x_3(t)x_4(t)$$

6) unità con legge di attivazione probabilistica. L'uscita, binaria, ha distribuzione di probabilità

$$Pr\{u(t+1)=y\} = 1/[1 + e^{-(P(t)-S)y/T}] \quad y = \pm 1$$

Per utilizzare queste unità occorre disporre di un generatore di numeri casuali con distribuzione uniforme in $[0, 1]$.

- Fra i tipi di **reti** maggiormente usati:

7) **reti prive di localizzazione**. Non si tiene conto della distanza tra le varie unità. Le unità sono contrassegnate da un indice. E' definita la matrice delle connessioni

$$W = [w_{ij}] \quad w_{ij} \text{ coeff. tra l'uscita dell'unità } j \text{ e l'ingresso dell'unità } i$$

Def. La rete è simmetrica $\Leftrightarrow W$ è simmetrica ($w_{ij} = w_{ji}$)

Def. La rete è antisimmetrica $\Leftrightarrow W$ è antisimmetrica ($w_{ij} = -w_{ji}$)

Def. Il **potenziale di attivazione** dell'unità i -esima è la somma pesata delle uscite delle N unità afferenti:

$$P_i(t) = \sum_{j=1..N} w_{ij} x_j(t)$$

8) **reti localizzate**. Si tiene conto della distanza d_{ij} tra le unità i -esima e j -esima

$$w_{ij} = w_{ij}^0 f(d_{ij})$$

Esempi:	esponenziale	$f(d_{ij}) = \exp(-d_{ij}/\lambda)$
	a cappello messicano	$f(d_{ij}) = [1 - (d_{ij}^2/2\sigma^2)]\exp(-d_{ij}^2/2\sigma^2)$

Oss. Se si scelgono $w_{ij}^0 > 0$, la funzione a cappello messicano trasforma in inibitorie le connessioni nell'intorno $\sigma/\sqrt{2} < |d_{ij}| < 2\sigma$.

9) **Reti con connessioni di ordine superiore al primo.** Sono le reti formate da unità $\Sigma\Pi$.

Def. L'**ordine** della rete, n , è il numero massimo di ingressi a moltiplicare che un'unità può ricevere

La descrizione usa le n **matrici generalizzate**

$$W_1 = [w_{ij}^{(1)}], W_2 = [w_{ijk}^{(2)}], \dots, W_n = [w_{ijl, \dots, jn}^{(n)}]$$

dove il primo indice indica l'unità ricevente, mentre gli altri n indici individuano le unità il prodotto delle cui uscite eccita l'unità in questione.

Si ha

$$P_i(t) = \sum_{j=1..N} w_{ij}^{(1)} x_j(t) + \sum_{j,k=1..N} w_{ijk}^{(2)} x_j(t) x_k(t) + \dots + \sum_{i,jl, \dots, jn=1..N} w_{ijl, \dots, jn}^{(n)} x_{jl}(t) \dots x_{jn}(t)$$

- Le architetture a **strati** (lineari, bi- o tri-dimensionali) si dividono in **monostrato** e **pluristrato**.
- Le **connessioni intra-strato** (o **connessioni laterali**) collegano unità di uno stesso strato, le **connessioni inter-strato** collegano unità appartenenti a strati differenti.
- Le connessioni interstrato si dividono in **feedforward** e **feedback** (o **ricorrenti**) a seconda della direzione.

COMPUTAZIONE SIMBOLICA E SUBSIMBOLICA

- La descrizione e comprensione dell'attività mentale si articola secondo due possibili impostazioni, una **dualistica**, l'altra **monistico-emergentista**.
- J.A:Fodor ha ravvisato i limiti dell'impostazione dualistica dato che non potendo concepire gli stati mentali come rappresentazioni di una classe finita di eventi fisici di stimolazione (molti concetti, p.es. "bontà", "bellezza", non sono riconducibili a una classe definibile di eventi-stimolo), è costretta infine a creare due livelli separati di realtà, quella fisica e quella degli stati mentali. Questo divario i dualisti neanche tentano di colmarlo, e si limitano a identificare lo stato mentale con una sequenza simbolica, suscettibile di un processo di calcolo che è l'attività mentale. Scopo della Psicologia sarebbe quello di individuare il programma, o l'algoritmo, che guida la computazione simbolica dei processi cognitivi.

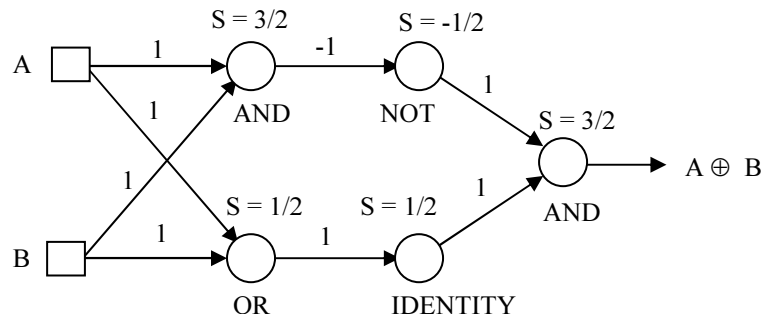
- L'impostazione monistico-emergentista ha indubbi vantaggi in efficienza, velocità, plausibilità biologica, ecc. Con semplici modelli a molte unità interconnesse si può fare in modo che l'evoluzione dei pattern di stimolazione porti il sistema a selezionare alcuni stati (gli **attrattori**) analoghi agli stati mentali. Ma l'analogia è debole, in quanto gli stati fisici non godono delle relazioni reciproche che la Psicologia e la Filosofia mostrano sussistere tra gli stati mentali. Questa difficoltà potrebbe essere superata immaginando architetture gerarchiche in cui gli strati dei livelli superiori abbiano attrattori via via più slegati dalle stimolazioni esterne.

RETI BINARIE DI McCULLOCH E PITTS

- Le **reti di McCulloch e Pitts** sono reti di unità binarie a soglia, prive di autoconnessioni. Nella versione iniziale, le singole unità di McCulloch e Pitts (i **neuroni formali**) avevano solo uno o due ingressi. Per una rete con un numero finito di unità di ingresso e una sola unità di uscita si potevano dimostrare i seguenti teoremi:

Th I. Qualunque unità booleana è equivalente a una rete di McCulloch-Pitts.

(P.es. un OR si ottiene con una sola unità binaria avente soglia $S = \frac{1}{2}$ e due ingressi con peso $w_1 = w_2 = 1$). L'or esclusivo (xor) è realizzato in tre passi dalla rete seguente



Il libro “*Perceptrons*” di *M.Minsky e S.Papert* (1969) dimostrava che non è possibile risolvere il problema dello xor *in un solo passo*. In realtà è possibile implementare qualunque funzione booleana, pur di avere a disposizione un tempo sufficientemente lungo.

Th II. La classe dei FA è equivalente alla classe delle reti di McCulloch-Pitts.

- *Def.* Una unità è di input se non riceve retroazioni dalle altre unità della rete. Una unità è di output se la sua uscita non è collegata alle altre unità.
- Talvolta si presentano reti ad attrattori dove tutte le unità sono sia di ingresso che di uscita. Ma in generale la rete ha r unità di ingresso, s unità di uscita e N unità nascoste. Il numero di stati di ingresso è 2^r , quello di uscita 2^s , mentre gli **stati interni** (o sempl. **stati**) sono 2^N .
- La legge di evoluzione della rete tra i suoi stati interni può essere

- a) **dinamica sincrona** (o **parallela**, o *di Little*) se tutte le unità aggiornano l'uscita negli stessi istanti. Dal punto di vista biologico è abbastanza inaccettabile.
- b) **dinamica asincrona** (o *di Hopfield*) se a ogni istante solo una unità scelta a caso aggiorna l'uscita (*asincrona casuale*). Un'alternativa è stabilire un ordine fisso con cui le unità vengono aggiornate (*asincrona sequenziale*).
- Supponiamo di avere un'unità di ingresso e un'unità di uscita collegate a ciascuna delle N unità nascoste. Così facendo in realtà la rete non è più un dispositivo di input-output. Input e output devono essere trattati in tempi diversi. Quello che era un trasduttore lineare è diventato un sistema dinamico complesso. Applichiamo inoltre le ipotesi:
 - 1) Coefficienti di connessione sia positivi che negativi
 - 2) Unità completamente connesse

In riferimento alla plausibilità neurofisiologica, l'ipotesi di connettività totale non è molto realistica. Tuttavia, considerato che la connettività media nella corteccia cerebrale è di $10^3 \div 10^4$ sinapsi per neurone, scegliendo reti con $N = 10^3 \div 10^4$ si può tollerare per il momento questa forzatura.

Applicato uno stato iniziale \mathbf{x}^0 in t_0 , l'**evoluzione libera** della rete (cioè con ingressi disabilitati) può essere di tre tipi:

 - a) **terminante in uno stato di equilibrio**, o **stato stazionario**, o **punto fisso** $\Leftrightarrow \exists t_{tr} \mid \text{per } t > t_{tr} \mathbf{x}(t) = \text{cost.}$
 - b) **terminante in un ciclo** $\Leftrightarrow \exists t_{tr}, \tau \mid \text{per } t > t_{tr} \mathbf{x}(t + \tau) = \mathbf{x}(t)$
 Def. τ è la **lunghezza** del ciclo
 Def. t_{tr} è la **lunghezza** del transitorio
 - c) **Caotico** \Leftrightarrow la sequenza $\mathbf{x}(t)$ è priva di ogni regolarità.
- Il problema principale è sapere sotto quali condizioni è possibile associare a uno stato iniziale una evoluzione libera tra i tipi citati
- Oss. Dal punto di vista neurofisiologico un punto fisso della rete corrisponderebbe a una situazione in cui mentre alcune unità sono attivate alla massima frequenza possibile, le altre sono completamente silenziose.
- Th III. L'evoluzione libera di una rete binaria sincrona termina sempre in un ciclo o in uno stato di equilibrio
 Dim. La rete ha un numero finito di stati, quindi uno stato prima o poi si ripresenta. Dato che la rete è deterministica, c.v.d.
 Oss. Questo teorema è simile al Pumping Lemma per i RLs. Anche il Pumping Lemma rivelava dei cicli, all'interno delle parole di un linguaggio. Ma nel caso dei FA l'evoluzione non è libera, bensì forzata dai simboli, che a ogni passo l'automa legge sul nastro in ingresso.
- Th IV. L'evoluzione libera di una rete binaria asincrona può essere indefinitamente caotica
- I ricercatori hanno cercato le condizioni cui deve sottostare una rete asincrona per avere stati di equilibrio

- *Th V.* Data una rete binaria asincrona e una funzione univoca $E(\mathbf{x})$ definita sui suoi stati, tale che

$$\forall t, E(\mathbf{x}(t+1)) \leq E(\mathbf{x}(t))$$

allora l'evoluzione libera della rete termina in uno stato di equilibrio, punto di minimo per E (minimo relativo o assoluto).

Dim. Il numero di stati è finito, e così pure i valori assunti da E ...

- *Def.* E è detta **energia** o **funzione di Ljapunov**.
- *Th VI.* Data una rete con unità binarie a soglia (sia $x=\pm 1$), asincrona simmetrica priva di autoconnessioni, con potenziali di attivazione mai nulli ($P_i(t)-S_i \neq 0$), ha un'evoluzione libera che termina in uno stato di equilibrio, e una possibile funzione di Ljapunov è

$$E(\mathbf{x}) = - \frac{1}{2} \sum_{i,j=1..N} w_{ij} x_i x_j + \sum_{k=1..N} S_k x_k$$

dove S_k è la soglia dell'unità k -esima.

Dim. Sia l'unità r -esima quella selezionata casualmente per l'aggiornamento dell'uscita al tempo t , quindi, usando il fatto che $w_{ji} = w_{ij}$ e $w_{ii} = 0$,

$$\begin{aligned} E(\mathbf{x}(t)) &= - x_r(t) \sum_{j \neq r} w_{rj} x_j - \frac{1}{2} \sum_{i,j \neq r} w_{ij} x_i(t) x_j(t) + S_r x_r(t) + \sum_{k \neq r} S_k x_k(t) \\ E(\mathbf{x}(t+1)) &= - x_r(t+1) \sum_{j \neq r} w_{rj} x_j - \frac{1}{2} \sum_{i,j \neq r} w_{ij} x_i(t) x_j(t) + S_r x_r(t+1) + \sum_{k \neq r} S_k x_k(t) \\ \Delta E(\mathbf{x}) &= - [x_r(t+1) - x_r(t)] [\sum_{j \neq r} w_{rj} x_j(t) - S_r] = - [x_r(t+1) - x_r(t)] [P_r(t) - S_r] \end{aligned}$$

Si vede che

per $x_r(t+1) = x_r(t)$ è $\Delta E(\mathbf{x}) = 0$;

per $x_r(t+1) = -1$ e $x_r(t) = 1$ dev'essere $P_r(t) - S_r < 0$, quindi $\Delta E(\mathbf{x}) < 0$;

per $x_r(t+1) = 1$ e $x_r(t) = -1$ dev'essere $P_r(t) - S_r > 0$, quindi $\Delta E(\mathbf{x}) < 0$.

c.v.d.

- *Th VII* (o **primo teorema di Goles**). Data una rete con unità binarie a soglia, sincrona simmetrica priva di autoconnessioni, con potenziali di attivazione mai nulli, ha un'evoluzione libera che termina in uno stato di equilibrio o in un ciclo di lunghezza 2. In particolare, se $S_k = 0$, una possibile funzione di Ljapunov è

$$E(\mathbf{x}) = - \sum_{i=1..N} \left| \sum_{j=1..N; j \neq i} w_{ij} x_j \right|$$

Dim. Si ha

$$\begin{aligned} E(\mathbf{x}(t)) &= - \sum_i \left[\sum_{j \neq i} w_{ij} x_j(t) \right] \cdot \text{sgn} \left[\sum_{j \neq i} w_{ij} x_j(t) \right] \\ E(\mathbf{x}(t)) &= - \sum_i \left[\sum_{j \neq i} w_{ij} x_j(t) x_i(t+1) \right] \\ \Delta E(\mathbf{x}) &= - \sum_i \left[x_i(t+2) - x_i(t) \right] \sum_{j \neq i} w_{ij} x_j(t+1) \end{aligned}$$

Per ogni contributo i di ha che:

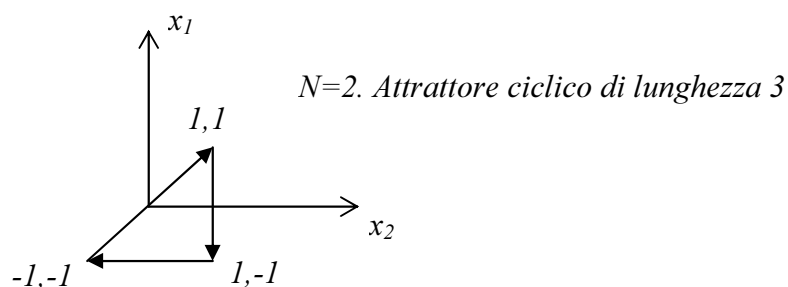
per $x_i(t+2) = x_i(t)$ è $\Delta E(\mathbf{x})=0$ (stato di equilibrio o ciclo limite di lunghezza 2)

per $x_i(t+2)=-1$ e $x_i(t)=1$ dev'essere $P_i(t+1)<0$, quindi $\Delta E(\mathbf{x})<0$;

per $x_i(t+2)=1$ e $x_i(t)=-1$ dev'essere $P_i(t+1)>0$,quindi $\Delta E(\mathbf{x})<0$.

c.v.d.

- *Th VII (o secondo teorema di Goles)*. Data una rete con unità binarie a soglia, sincrona antisimmetrica, con potenziali di attivazione mai nulli, ha un'evoluzione libera che termina sempre in un ciclo di lunghezza 4.
- Tra gli aspetti che restano da chiarire, la durata dei transitori, la relazione tra stato iniziale e evoluzione libera, lunghezza media dei cicli.



- In generale non è detto che la funzione di energia esista e non c'è un metodo generale per ottenerla in forma analitica, a meno che non sussistano ipotesi specifiche come nel caso dei precedenti teoremi.
- Per usare una rete ad attrattori come sistema cognitivo, capace di classificare, riconoscere o richiamare un'informazione in tempo reale, occorre tenere conto dei vincoli temporali imposti dall'ambiente. Il sistema deve rispondere in tempo utile agli stimoli che l'ambiente fornisce con una certa frequenza. La dinamica di convergenza della rete su un attrattore dev'essere quindi abbastanza veloce e inoltre il sistema di risposta comportamentale deve accorgersi rapidamente se la rete ha raggiunto un attrattore. Un limite superiore alla latenza totale potrebbe essere 40 ms (Stenberg, 1966; Stenberg, 1969). Questo potrebbe significare ad esempio che i bacini di attrazione utili saranno dei sottoinsiemi di quelli asintotici.

LE RETI BOOLEANE

- *Def.* La **connettività** K è il numero di ingressi delle unità, di solito inferiore al numero totale di unità N (diversamente dalle reti di McCulloch-Pitts).
- Di solito la topologia delle reti booleane a connettività 2 è piana toroidale (le unità e le connessioni si dispongono senza intersecarsi su una superficie chiusa).
- Tipicamente le funzioni implementate dalle singole unità $B: \{0,1\}^K \rightarrow \{0,1\}$ (in tutto 2^S , $S=2^K$) sono scelte a caso all'inizio, e poi mantenute per tutta la durata dell'esperimento.

- Fu **S.A.Kauffman** il primo a simulare le reti booleane sincrone (1969). Egli definì **attrattore** a ogni ciclo limite o stato di equilibrio, e il relativo **bacino di attrazione** A l'insieme degli stati iniziali che portano l'evoluzione a quell'attrattore. Egli poté calcolare la **dimensione** media dei bacini di attrazione $\langle |A| \rangle$, e la loro **forma** nello spazio degli stati basandosi su una **metrica di Hamming**

$$H(x,y) \equiv \sum_{i=1..N} x_i \oplus y_i = \sum_{i=1..N} |x_i - y_i|$$

- Oss. $H(x,y)$ è pari al numero di componenti per le quali x differisce da y
- Def. La **complessità** di una rete è

$$C = -\sum_{q=1..Q} (|A_q|/2^N) \log(|A_q|/2^N)$$

dove Q è il numero di attrattori.

Oss. Espressione identica all'informazione media H di Shannon.

- Def. Uno stato di equilibrio è **instabile** \Leftrightarrow cambiando anche una sola delle sue componenti, l'evoluzione libera della rete a partire da questo stato modificato non termina nello stato di equilibrio considerato.
- Per le reti a connettività totale ($K=N$) Kauffman calcolò la lunghezza media dei cicli, pari a $E\{\tau\} = 2^{N/2-1}$; il numero medio di attrattori, $E\{Q\} = N/e$;
- Per le reti a connettività $K=2$ Kauffman trovò $E\{\tau\} = E\{Q\} = N^{1/2}$, un risultato molto diverso da quello ottenuto già per $K>4$:

$$E\{\tau\} = 1/2 \{ 1/2 + 2^{-(2^K+1)} (2^{2^K} {}_{2^{(K-1)}}C_{2^{(K-1)}}) \}^{-N/2}$$

- Per spiegare questo comportamento suddivise le unità in tre categorie a seconda del loro comportamento dopo che la rete è entrata in un ciclo qualsiasi:
 - a) unità **stabili**, che mantengono l'uscita costante, $\forall x^0$
 - b) unità **oscillanti**, con uscita periodica, $\forall x^0$
 - c) unità **incerte**, con uscita costante o periodica a seconda di x^0 .
- Le reti con $K=2$ sono caratterizzate da pochi attrattori, con piccoli bacini stabili, separati da larghe regioni di unità stabili. Il motivo di questo è nel fatto che la percentuale di funzioni booleane aventi variabili di ingresso **forzanti** (che determinano l'uscita indipendentemente dalle altre variabili di ingresso) decresce rapidamente col numero di ingressi K . (75% per $K=2$, 5% per $K=4$).
- Def. Una connessione $i-j$ è **forzante** se $x_i(t)$ è un valore forzante applicato a un ingresso forzante dell'unità j . Un **circuito forzante** è un cammino chiuso di connessioni forzanti.
- Def. Il **dominio forzante** (*forcing domain*) D di una rete è l'insieme delle unità che
 - a) si trovano su un circuito forzante
 - b) riceve una connessione forzante da una unità in D
 - c) riceve tutti i suoi ingressi da unità in D
 - d) invia una connessione forzante a una unità in D

- *Th VIII*. Data una rete booleana sincrona di connettività 2, per ogni circuito forzante C vale una delle due condizioni
 - i. Le unità di C sono stabili
 - ii. Le unità di C sono oscillanti, con valori di ingresso sempre diversi dai valori forzanti
- *Th IX*. Data una rete booleana asincrona di connettività 2, per ogni circuito forzante C vale una delle due condizioni
 - i. Le unità di C sono stabili
 - ii. Le unità di C sono oscillanti
- In pratica si osserva che molto spesso il nucleo stabile e il dominio forzante di una rete coincidono
- **H.Atlan** applicava, in una rete con $K=2$, una sequenza casuale al posto dell'uscita di una unità incerta, per un tempo pari a 2τ . Ciò poteva portare una delle unità oscillanti in uno stato stabile. E' come se quell'unità avesse "riconosciuto" la sequenza applicata.
- Secondo **C.Langton** le reti a connettività non troppo elevata ma maggiore di 2 avrebbero le capacità maggiori di reagire a stimoli esterni, cioè in modo né troppo caotico né troppo prevedibile.

RETI DI CAIANIELLO

- Def. Una **rete di Caianiello** (1961) è una generalizzazione della rete binaria di McCulloch-Pitts in cui il potenziale di attivazione di ogni unità dipende non solo dai valori degli ingressi nell'istante considerato, ma anche dai valori assunti in istanti precedenti, cioè

$$x_i(t+1) = \Psi(P_i(t) - S_i)$$

$$P_i(t) = \sum_{j=1..N} \sum_{r=0..L} w_{ij}^{(r)} x_j(t-r)$$

- Le equazioni sopra riportate furono chiamate da Caianiello **equazioni neuroniche**, accanto alle quali egli pose le **equazioni mnemoniche**, che descrivono la dinamica dei coefficienti $w_{ij}^{(r)}$ nei processi di apprendimento, e che evolverebbero con costanti di tempo più lunghe rispetto alle equazioni neuroniche (**ipotesi dell'apprendimento adiabatico**).
- **Esempio**. Una rete di Caianiello con $N=2$ e $L=1$ ha equazioni di partenza

$$x_1(t+1) = \Psi(w_{11}^{(0)} x_1(t) + w_{12}^{(0)} x_2(t) + w_{11}^{(1)} x_1(t-1) + w_{12}^{(1)} x_2(t-1) - S_1)$$

$$x_2(t+1) = \Psi(w_{21}^{(0)} x_1(t) + w_{22}^{(0)} x_2(t) + w_{21}^{(1)} x_1(t-1) + w_{22}^{(1)} x_2(t-1) - S_2)$$

- Il primo passo è quello di eliminare la dipendenza dagli istanti precedenti a t , introducendo NL variabili ausiliarie (unità fittizie) con equazioni

$$x_k(t+1) = \Psi(x_{k-N}(t) - 1/2), \quad k = N+1, \dots, N+NL$$

Se l' i -esima equazione verifica

$$S_i = 1/2 \sum_{j=1..N} w_{ij}, \quad i=1, \dots, N$$

allora il suo termine noto è nullo, altrimenti si aggiunge una nuova variabile ausiliaria con equazione

$$x_k(t+1) = \Psi(x_k(t) - 1/2)$$

Si opera poi la sostituzione di variabili $x = 1/2(\xi+1)$, ottenendo un primo sistema, in **forma normale**, di M equazioni algebriche in altrettante incognite $\xi_i(t)$, con coefficienti w_{ij} ,

$$\xi_i(t+1) = \text{sgn}[1/2 \sum_{j=1..M} w_{ij} \xi_j(t)] \quad N(L+1) \leq M \leq N(L+2),$$

Notiamo che mentre $x \in \{0, 1\}$, si ha $\xi \in \{-1, 1\}$.

- Per rimuovere l'ostacolo della non invertibilità delle funzioni $\text{sgn}()$ si ricorre al procedimento di linearizzazione tensoriale, secondo cui data una funzione $F(\xi_1, \dots, \xi_M)$ mai nulla di M variabili booleane, $\text{sgn}(F)$ è esprimibile come combinazione lineare a coefficienti costanti dei prodotti tensoriali $\eta_k \equiv \xi_1^{k_1} \dots \xi_M^{k_M}$, con $k_M \dots k_1$ codifica binaria dei numeri $k = 1, \dots, 2^M$.

$$\text{sgn}[F(\xi_1, \dots, \xi_M)] = \sum_k c_k \eta_k$$

I coefficienti sono dati dalla formula

$$c_k = \langle \eta_k \text{sgn}[F(\xi_1, \dots, \xi_M)] \rangle$$

dove si è definito l'operatore

$$\langle B(\xi_1, \dots, \xi_M) \rangle = (1/2)^M \sum_{\xi_i = \pm 1} B(\xi_1, \dots, \xi_M)$$

essendo B una funzione booleana.

- In definitiva si è potuto trasformare il sistema in forma normale nel nuovo sistema di 2^M equazioni nelle altrettante incognite $\eta_k(t)$,

$$\eta_k(t+1) = \sum_{r=1..2^M} F_{kr} \eta_r(t), \quad k = 1, \dots, 2^M$$

- *Th. X.* L'equazione caratteristica della matrice F (2^M righe \times 2^M colonne) è del tipo

$$\lambda^a \prod_b (\lambda^b - 1)^{C_b} = 0, \quad \text{con } C_b \geq 0 \quad \text{e} \quad a + \sum_b b \cdot C_b = 2^M$$

- *Th. XI.* Se $a > 0$ (autovalore $\lambda = 0$) l'evoluzione libera della rete ha un transitorio. Inoltre, se si hanno valori C_b non nulli, ciascun termine $(\lambda^b - 1)^{C_b}$ denota l'esistenza di C_b cicli limite di periodo b .

TECNICHE DI OSSERVAZIONE SPERIMENTALE

- Cosa è possibile e cosa conviene osservare in una simulazione di una rete binaria?
- Si possono considerare aspetti noti ($E\{\tau\}$, $\max\{\tau\}$, $E\{t_{tr}\}$, $\max\{t_{tr}\}$, $E\{Q\}$, $E\{|A|\}$, C , stabilità dei bacini di attrazione e stabilità media, ecc.) oppure cercare una regolarità nelle caratteristiche temporali o spaziali, non note a priori.
- Per determinare sperimentalmente la lunghezza media di un ciclo, $E\{\tau\}$, bisogna eseguire un certo numero di esperimenti e per ciascuno memorizzare le sequenze

$$x_i(t) \quad t = t_0, \dots, t_f - 1, t_f \quad i = 1, \dots, N$$

Per ciascun esperimento, τ è dato dalla condizione $\mathbf{x}(t_f - \tau) = \mathbf{x}(t_f)$.

- Per calcolare $E\{\tau\}$, essendo praticamente impossibile determinare τ con una simulazione esaustiva delle evoluzioni a partire da tutti gli stati iniziali, si deve procedere alla verifica statistica della significatività dei risultati ottenuti su un campione di stati iniziali. Per determinare la lunghezza del transitorio si applica il seguente algoritmo:

```
for(r=1;  $\mathbf{x}[t_f - (r+1)\tau] \neq \mathbf{x}[t_f]$ ; r++);
for(k=1;  $\mathbf{x}[t_f - (r-1)\tau + k] \neq \mathbf{x}[t_f - r\tau + k]$ ; k++);
return( $t_f - r\tau + k - t_0$ );
```

- P.es. trovato che in n esperimenti si misurano m classi $\tau = \tau_i$ con frequenza relativa f_i . Si ipotizza una distribuzione gaussiana, con media $\mu = \sum_i \tau_i f_i$ e varianza $\sigma^2 = \sum_i (\tau_i - \mu)^2 f_i$. Poi si calcolano le frequenze relative teoriche $p_i = (1/\sigma\sqrt{2\pi}) \exp[-(\tau_i - \mu)^2 / 2\sigma^2]$. Infine si calcola $\chi^2 = n \sum_i (\tau_i - p_i)^2 / p_i$ che va confrontato con la tabella della distribuzione teorica di χ^2 con un numero di gradi di libertà pari a $m-1-2$ (il -2 sta per il numero di vincoli imposti alla distribuzione, cioè μ e σ) e un livello di significatività scelto (p.es. 0.05, cioè fiducia al 95%). Se il χ^2 è inferiore al valore teorico possiamo accettare l'ipotesi, altrimenti dobbiamo fare altri esperimenti.
- *Def.* Altri parametri di interesse sono:
l'**attività media**

$$\rho(t) = (1/N) \sum_{i=1..N} x_i(t)$$

la **varianza media**

$$\sigma^2(t) = (1/N) \sum_{i=1..N} [x_i(t) - x_i(t_0)]^2$$

la **distanza media di Hamming normalizzata tra stati consecutivi**

$$H(t) = (1/N) \sum_{i=1..N} |x_i(t) - x_i(t-1)| = (1/N) H(\mathbf{x}(t), \mathbf{x}(t-1))$$

Il **coefficiente di autocorrelazione temporale medio** in $[t_0, t_0+T]$

$$C(\theta) = (1/N) \sum_{i=1..N} \{ \sum_{t=t_0+1..t_0+T} x_i(t)x_i(t+\theta)/T - [\sum_{t=t_0+1..t_0+T} x_i(t)/T]^2 \}$$

Il grafico di $r(\theta) \equiv C(\theta)/C(0)$ è detto **correlogramma**. Ricordiamo che valgono le seguenti approssimazioni

$$E\{r(\theta)\} \approx -1/T$$

$$E\{[r(\theta) - E\{r(\theta)\}]^2\} \approx 1/T$$

quindi, a un livello di significatività di 0.05, i valori di r da considerare attendibili sono quelli per cui $|r(\theta) + 1/T| > 2/\sqrt{T}$.

Se $\forall \theta, |r(\theta) + 1/T| \approx 2/\sqrt{T}$ la rete ha una evoluzione temporale caotica. Al contrario se $|r(\theta) + 1/T| \gg 2/\sqrt{T}$ la rete ha una tendenza sistematica. Se esiste una oscillazione periodica allora anche la rete ha evoluzioni cicliche.

L'**intensità media della p-esima armonica di Fourier**, p intero in $[1, T/2]$,

$$I(p) = (1/NT) \sum_{i=1..N} \{ [\sum_{t=t_0+1..t_0+T} x_i(t) \cos(2\pi p t/T)]^2 + [\sum_{t=t_0+1..t_0+T} x_i(t) \sin(2\pi p t/T)]^2 \}$$

Picchi nel grafico di $I(p)$ vs. p (periodogramma) denotano componenti periodiche di periodo $\tau = T/p$.

la **funzione di correlazione spaziale**, che per una rete unidimensionale è

$$G(k) = (1/T) \sum_{t=t_0+1..t_0+T} \{ \sum_{i=1..N} x_i(t)x_{i+k}(t)/N - [\sum_{i=1..N} x_i(t)/N]^2 \}$$

Anche in questo caso i valori da prendere in considerazione sono quelli per cui $|G(k) + 1/N| > 2/\sqrt{N}$.

Lo **spettro di potenza spaziale**,

$$S(k) = (1/T) \sum_{t=t_0+1..t_0+T} |\sum_{i=1..N} x_i(t) \exp(j2\pi k i)|$$

RETI AD APPRENDIMENTO DIRETTO

- Una **memoria associativa** è un sistema capace di immagazzinare dei pattern di informazione e, una volta stimolato con un pattern di ingresso (*probe*), emettere uno dei pattern memorizzati. Di solito si tratta di **content addressable memory (CAM)**, come nel modello di Hintzman. Distinguiamo memorie **autoassociative** (il *probe* è una copia deteriorata di uno dei pattern immagazzinati, e il sistema deve recuperare il pattern prototipo) e memorie **eteroassociative**.

IL MODELLO DI HOPFIELD

- Una rete di Hopfield è totalmente connessa, simmetrica e priva di autoconnessioni. Le unità sono binarie ($x = \pm 1$) a soglia.
- La nota **legge di Hebb** viene formalizzata ammettendo una variazione dell'efficacia sinaptica Δw_{ij} di una connessione tra l'uscita dell'unità j -esima e l'ingresso dell'unità i -esima proporzionale al prodotto di delle attività $x_j(t)$ e $x_i(t+s)$, dove s è il ritardo di propagazione (trascurabile in prima approssimazione).
- Durante la **fase di memorizzazione** si forza l'uscita delle N unità ai valori del primo pattern di ingresso (il primo item da memorizzare), un vettore \mathbf{x}^1 di N valori ± 1 . Ne risulta un incremento per il generico coefficiente di connessione

$$\Delta w_{ij} = \eta x_i x_j$$

Continuando allo stesso modo con i successivi pattern (in tutto R , anche con ripetizioni), e scegliendo $\eta = 1/N$, si arriva a un incremento cumulativo (inizialmente $w_{ij} = 0$)

$$w_{ij} = (1/N) \sum_{s=1..R} x_i^s x_j^s - \delta_{ij}/N$$

- Oss. il termine δ_{ij}/N serve ad annullare gli elementi sulla diagonale principale di \mathbf{W} , nel rispetto del requisito di assenza di autoconnessioni.
- Oss. Potremmo anche scrivere sinteticamente la matrice delle connessioni al termine della fase di apprendimento (R item in tutto) definendo la matrice $N \times R$ dei pattern di ingresso $\mathbf{X} = [\mathbf{x}^1 \ \mathbf{x}^2 \ \dots \ \mathbf{x}^R]$

$$\mathbf{W} = (1/N) [\mathbf{X} \cdot \mathbf{X}^T - \mathbf{R}\mathbf{I}]$$

- Nella **fase di richiamo** si forza per un istante un pattern di stimolo sulle uscite delle N unità, quindi si lascia la rete in evoluzione libera (di solito asincrona), tenendo i coefficienti di connessione bloccati ai valori risultanti dalla fase di apprendimento. Per il *Th. VI* la rete raggiungerà uno stato di equilibrio, che viene

identificato col pattern richiamato. Successivamente si può procedere a piacere con nuove fasi di richiamo e memorizzazione.

- Supponendo di usare come pattern di richiamo proprio uno dei pattern memorizzati, cioè $\mathbf{x}(t_0) = \mathbf{x}^q$, nell'ipotesi semplificativa $S_i = 0$, avremo

$$\begin{aligned} x_i(t_0+1) &= \text{sgn}[\sum_{j \neq i} w_{ij} x_j^q] = (1/N) \text{sgn}[\sum_{j \neq i} \sum_{s=1..R} x_i^s x_j^s x_j^q] = \\ &= \text{sgn}[x_i^q + (1/N) \sum_{s \neq q} x_i^s \sum_{j \neq i} x_j^s x_j^q] \end{aligned}$$

Il primo termine tra parentesi favorisce l'allineamento di x_i nella direzione di x_i^q , che è proprio quella del pattern che dev'essere richiamato. Questo termine rappresenta il segnale utile, mentre il secondo termine rappresenta un rumore. In particolare, per N e R abbastanza grandi si ha

$$\sum_{s \neq q} x_i^s \sum_{j \neq i} x_j^s x_j^q = O[(RN)^{1/2}]$$

sicché

$$x_i(t_0+1) = \text{sgn}\{x_i^q + O[(R/N)^{1/2}]\}$$

Allora, se $R \ll N$, $\mathbf{x}(t)$ tenderà a coincidere con \mathbf{x}^q per $t > t_0$. Lo stesso discorso vale qualora $\mathbf{x}(t_0)$ e \mathbf{x}^q differiscano di poco.

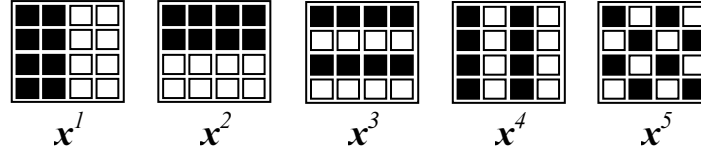
- Se però $\mathbf{x}(t_0)$ non coincide con uno dei pattern memorizzati, la rete di Hopfield non dà garanzie assolute che il pattern richiamato coincida con uno dei pattern memorizzati. P.es. se ammettiamo sempre $P_i \neq 0$, e se immagazziniamo un solo pattern \mathbf{x}^l , avremo che oltre a \mathbf{x}^l anche $-\mathbf{x}^l$ è uno stato di equilibrio, che potrebbe essere sempre richiamato, e che non coincide con il pattern memorizzato. Questo fatto discende immediatamente dall'invarianza dei coefficienti w_{ij} rispetto alla trasformazione $\mathbf{x}^s \rightarrow -\mathbf{x}^s$. Se poi ammettiamo anche $P_i = 0$, pure $\mathbf{x} = (1 \ 1 \dots 1)^T$ è uno stato di equilibrio.
- Def. La **sovrapposizione** (*overlap*) tra due stati \mathbf{x} e \mathbf{y} è

$$C(\mathbf{x}, \mathbf{y}) = (1/N) \sum_{i=1..N} x_i y_i$$

- Oss. $C(\mathbf{x}, \mathbf{y}) \equiv (1/N) \mathbf{x} \cdot \mathbf{y} \equiv 1 - 2H(\mathbf{x}, \mathbf{y})/N$
- Sono stati ottenuti dei risultati importanti sul comportamento delle reti di Hopfield usando i metodi della Meccanica Statistica e del Calcolo delle Probabilità, nelle ipotesi semplificative:
 - a) $N \gg 1$
 - b) $R \ll N$
 - c) I pattern da memorizzare sono scelti a caso. Ciò implica che la media delle componenti di un dato pattern, $|(1/N) \sum_{i=1..N} x_i^s|$, è minore o uguale a $1/\sqrt{N}$ (dalla Meccanica Statistica è noto che le fluttuazioni spontanee delle

grandezze globali hanno ampiezza massima $1/\sqrt{N}$). Inoltre questi pattern sono anche scorrelati, cioè hanno una sovrapposizione $|C(\mathbf{x}^s, \mathbf{x}^q)| \leq 1/\sqrt{N}$, $s \neq q$.

Esempio. La figura mostra un esempio di $R=5$ pattern che soddisfano le condizioni di cui sopra per una rete di $N=16$ unità, avendo media e sovrapposizioni mutue nulle.

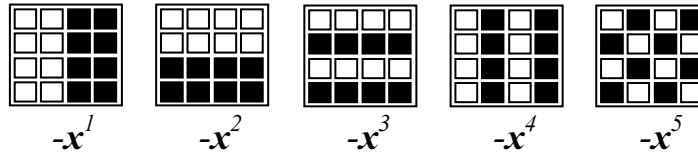


- *Th.* I pattern $\pm \mathbf{x}^1, \dots, \pm \mathbf{x}^R$ sono tutti stati di equilibrio stabile (*attrattori puri*), sia nel caso di evoluzione sincrona che asincrona. La funzione di Ljapunov

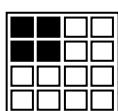
$$E(\mathbf{x}) = - \frac{1}{2} \sum_{i \neq j} w_{ij} x_i x_j + \sum_{k=1..N} S_k x_k$$

assume minimi assoluti in questi punti.

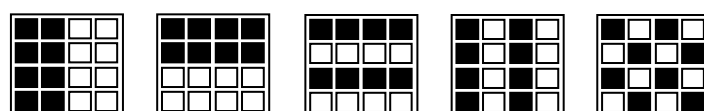
Esempio. Nell'esempio precedente avremmo anche i seguenti attrattori puri



- *Th.* Esistono altri stati di equilibrio detti **attrattori spuri**. Questi stati hanno overlap non nulli con alcuni dei pattern memorizzati. Gli attrattori spuri hanno un contenuto energetico superiore a quello degli *attrattori puri*, e quindi sono minimi relativi per E , cioè *stati metastabili*.
- *Def.* Uno stato spurio è una **mistura simmetrica** \Leftrightarrow ha stesso overlap con due o più pattern memorizzati, e overlap nullo con tutti gli altri.
- *Th.* La rete ha 3^R stati misture.
- *Th.* Esistono 2^k ($k \leq R$) stati spuri che sono misture simmetriche di k pattern memorizzati.
- *Th.* Gli stati misture simmetriche di un numero pari di pattern memorizzati sono punti di equilibrio instabile per la dinamica sequenziale e appartengono a cicli di lunghezza 2 per la dinamica parallela. Gli stati misture simmetriche di un numero dispari di pattern memorizzati sono sempre punti di equilibrio stabile, cioè sono attrattori.
- **Esempio.** Lo stato \mathbf{y} è una mistura simmetrica di un numero pari di pattern memorizzati



\mathbf{y}



$$C(\mathbf{y}, \mathbf{x}^1) = \frac{1}{2} \quad C(\mathbf{y}, \mathbf{x}^2) = \frac{1}{2} \quad C(\mathbf{y}, \mathbf{x}^3) = 0 \quad C(\mathbf{y}, \mathbf{x}^4) = 0 \quad C(\mathbf{y}, \mathbf{x}^5) = 0$$

- *Th.* Possono esistere punti di equilibrio che hanno sovrapposizione diverse con due o più pattern memorizzati (**attrattori spuri asimmetrici**)
- *Th.* $\exists R_c \mid \forall R > R_c$ gli stati $\pm x^1, \dots, \pm x^R$ cessano di essere punti di equilibrio stabili. In particolare, se si accettano scostamenti $< 1\%$ degli stati di equilibrio dai pattern memorizzati allora questi stati di equilibrio sono stabili per

$$R < R_c \equiv \alpha_c N = 0.145N$$

Se $R > R_c$ il richiamo di questi pattern è impossibile (**catastrofe di blackout**).

Questo significa che per immagazzinare 10 pattern dobbiamo disporre di almeno $R/\alpha_c = 68.9 \Rightarrow 70$ unità.

- Il modello di Hopfield ha forti limiti, sia come modello psicologico della memoria che ai fini delle applicazioni pratiche:
 - a) ha scarse capacità di immagazzinamento
 - b) le stime a priori delle capacità della rete sono possibili sotto l'ipotesi non realistica di pattern a media nulla e mutuamente scorrelati
 - c) la condizione di non correlazione è anche quella che massimizza R_c . Si ricava che per una sovrapposizione costante $|C| > 1/\sqrt{N}$ tra tutti i pattern immagazzinati, è $R_c = 1 + 1/C$. P.es. con $N = 100$ unità possiamo memorizzare fino a $R_c = 14$ pattern con correlazione mutua $< 1/\sqrt{N} = 0.1$. Se invece i pattern avessero correlazione mutua $C = 0.2$, potremmo memorizzarne al più $R_c = 1 + 1/0.2 = 6$.
 - d) un altro problema sono gli attrattori spuri, di cui non è ancora possibile predire l'ampiezza dei bacini di attrazione
 - e) è difficile interpretare tutti quei dati sperimentali sulla memoria umana che implicano l'intervallo di latenza, dato che non è possibile conoscere a priori la durata della fase di richiamo
 - f) il vincolo della simmetria delle connessioni appare troppo poco realistico
- Il pregio del modello è aver dimostrato la realizzabilità di una *CAM*, che richiama alcune caratteristiche del funzionamento del sistema nervoso, e che è robusta rispetto al malfunzionamento (simulabile con una legge di attivazione probabilistica) o alla eliminazione casuale delle connessioni (se c è la probabilità di eliminazione di una connessione, si dimostra che $R_c = (1-c)\alpha_c N$).
- I modelli deterministici trascurano l'aleatorietà intrinseca dei processi nervosi. Varie sorgenti di rumore sono state sperimentalmente identificate nella trasmissione sinaptica (Katz, 1966)
 - 1) il numero di vescicole che va incontro a esocitosi all'arrivo del potenziale d'azione varia in modo casuale con una distribuzione di probabilità di Poisson.

Il coefficiente di efficacia sinaptica w_{ij} rappresenterebbe proprio il valore medio di questa distribuzione

- 2) Il contributo al potenziale post-sinaptico dato da ciascun quanto di neurotrasmettitore (un quanto sarebbe il contenuto di una vescicola) ha distribuzione normale. Si assume tale distribuzione uguale per tutte le sinapsi della rete
 - 3) Anche in assenza di potenziale d'azione esiste un tasso casuale di rilascio spontaneo di neurotrasmettitore
- Tenendo conto di questi effetti, il potenziale di attivazione diventa una v.a. P_i con valore medio $\langle P_i \rangle = P_i$ e distribuzione gaussiana

$$p(P) = (1/2\pi\sigma^2)^{1/2} \exp [-(P-P_i)^2/2\sigma^2]$$

essendo σ legato ai parametri associati alle diverse sorgenti di rumore, mentre $\langle P_i \rangle$ è il potenziale medio, che evolve secondo l'equazione solita

$$\langle P_i \rangle = P_i(t) = \sum_{j=1..N} w_{ij} x_j(t)$$

La probabilità che il neurone scarichi alla massima frequenza ($x_i = 1$) è pari alla probabilità che sia $P_i(t) > S_i$, cioè

$$Pr\{x_i(t+1) = 1\} = \int_{S_i, +\infty} p(P) dp = 1/2 \{1 + \text{erf}[(P_i(t) - S_i)/\sigma\sqrt{2}]\}$$

Accettando un'approssimazione all'1% si ricava un'espressione per la legge di attivazione che permette di stabilire una analogia con la meccanica statistica dei sistemi di Ising

$$Pr\{x_i(t+1) = s\} = 1/2 \{1 + \tanh [\beta(P_i(t) - S_i)s]\} = 1/\{1 + \exp[-2\beta(P_i(t) - S_i)s]\} \quad s=\pm 1$$

dove $T \equiv 1/\beta = \sigma\sqrt{2}$ è la **temperatura** della rete. Questa espressione è infatti identica a quella che descrive la dinamica di un singolo spin (dinamica di Glauber) in un sistema di Ising a temperatura T . L'introduzione di una temperatura non nulla nella rete di Hopfield equivale in definitiva all'introduzione di rumore. Se $T \rightarrow 0$ la probabilità vale 1, e valgono i risultati delle reti deterministiche. Nel caso probabilistico si complica la definizione di punto di equilibrio. Inoltre non ha più senso parlare di stati stabili ma di distribuzione degli attrattori.

- *Def.* Uno stato a è di equilibrio per una dinamica probabilistica del tipo specificato \Leftrightarrow la sovrapposizione tra $x(t)$ e a è un processo stocastico con media 1 e deviazione standard $< 1/\sqrt{N}$.
- *Th.* $\exists T_c = 1$ (**temperatura critica**) | $\forall T > T_c$ il comportamento della rete è casuale, o *ergodico*, cioè non esistono attrattori.

- Th. per $0.461 < T < T_c$ gli unici stati di equilibrio sono gli stati puri, cioè i pattern memorizzati e quelli complementari $\pm \mathbf{x}^1, \dots, \pm \mathbf{x}^R$. E' importante notare che in questo intervallo non ci sono attrattori spuri. Per $T < 0.461$, al diminuire di T , cominciano a divenire stati stabili le misture simmetriche di 3 pattern memorizzati, poi quelli di 5, e così via. Gli stati puri restano però sempre gli stati di minimo assoluto per l'energia E , fino alla temperatura $T = 0$.
- L'introduzione di una giusta quantità di rumore nella dinamica di una rete ad attrattori riduce la probabilità di cadere negli stati spuri della rete, rappresentati da minimi locali della funzione di energia (se esiste), migliorando quindi le prestazioni di richiamo delle informazioni. Probabilmente questo ruolo positivo del rumore è una caratteristica universale dei sistemi viventi (Atlan, 1972). I sistemi biologici sono estremamente complessi e come tali ricchi di stati metastabili, la maggior parte dei quali risulterebbe probabilmente mortale se il sistema dovesse persistervi.
- Oss. Lo stato di una rete di Hopfield probabilistica è dato dalle N v.a. $x_i(t)$. In dinamica sincrona il processo multivariato $\mathbf{x} = (x_1, \dots, x_N)$ è **un processo di Markov del primo ordine**: la specificazione di tutte le variabili in t è sufficiente a predire in senso statistico lo stato del sistema nell'istante successivo $t+1$. La probabilità di transizione è

$$W(\mathbf{x}|\mathbf{x}^0) = Pr(\mathbf{x}=\mathbf{x}, t+1 | \mathbf{x}=\mathbf{x}^0, t) = \frac{1}{2} \prod_{i=1 \dots N} \{1 + \tanh [\beta (\sum_{j=1 \dots N} w_{ij} x_j^0 - S_i) x_i]\}$$

GENERALIZZAZIONI DEL MODELLO DI HOPFIELD

- Le generalizzazioni della rete di Hopfield riguardano:
 - a) la regola di memorizzazione
 - b) la dinamica nella fase di richiamo
 - c) il ruolo delle soglie
- Nei **modelli nonlineari**, calcolata la matrice delle connessioni sulla base degli R pattern da memorizzare,

$$w_{ij} = (1/N) \sum_{s=1 \dots R} x_i^s x_j^s$$

la si fa passare per una trasformazione nonlineare

$$J_{ij} = \phi(w_{ij})$$

- **Van Hemmen** ha dimostrato che in queste reti oltre agli stati stabili $\pm \mathbf{x}^1, \dots, \pm \mathbf{x}^R$ ve ne sono altri di equilibrio stabile (o metastabile) formati dal prodotto di due o più vettori $\pm \mathbf{x}^1, \dots, \pm \mathbf{x}^R$, dove qui per prodotto si intende $p_i = x_{i_1}^s x_{i_2}^p, i = 1, \dots, N$.

- Un caso particolare di interesse è quello delle **sinapsi tagliate** (*clipped synapses*), dove $J_{ij} = (\sqrt{R/N}) \operatorname{sgn}(w_{ij})$. Purtroppo questo modello peggiora le capacità di memoria rispetto al modello lineare di Hopfield, portando α_c da 0.145 a 0.1.
- Per evitare il sovraccarico della memoria per $R > R_c$, **Nadal, Toulouse, Changeux e Dehaene** proposero la variante dell'**apprendimento limitato** (*learning with bounds*), che permette alla rete di dimenticare i dati appresi nel passato a favore di quelli più recenti. La matrice delle connessioni si modifica in fase di memorizzazione del pattern s -esimo in base alla

$$J_{ij}^1 = (1/N) x_i^1 x_j^1$$

$$J_{ij}^s = f[J_{ij}^{s-1} + C x_i^s x_j^s] \quad \text{per } s > 1$$

dove $f(x) = x$ per $|x| < A$; $A \cdot \operatorname{sgn}(x)$ altrimenti. Tipicamente $A = 1/\sqrt{N}$. Se $C \ll A$ il modello è simile a quello di Hopfield. Se invece $C \gg A$ l'ultimo pattern è memorizzato a spese del primo che viene cancellato. Si ricava $\alpha_c = 0.016$. Questo modello ha le migliori prestazioni nella rievocazione degli ultimi pattern memorizzati, ma le peggiori per i primi. In questo non rispecchia le caratteristiche della memoria umana, dove esiste un effetto *primacy*.

- La generalizzazione di **Personnaz, Guyon e Dreyfus** fissa i valori dei coefficienti a

$$W = (1/N) X \cdot X^T$$

dove si assume X formato da pattern ottenuti per combinazione lineare di r vettori linearmente indipendenti, con $r < N$. X^T è la **pseudoinversa** di X , $X^T \equiv (X^T \cdot X)^{-1} \cdot X^T$. Un aspetto è che l'evoluzione sincrona termina sempre in uno stato stabile.

- Le **reti asimmetriche** hanno maggiore aderenza ai dati neurofisiologici. Poco si sa sulla loro dinamica ma di certo l'asimmetria tende a ridurre la probabilità di richiamo di stati spuri. Nel caso della **asymmetric dilution** si annulla a caso una delle due connessioni simmetriche su una certa percentuale g del totale, $J_{ij} = w_{ij} \Psi(h - 1/2)$, $J_{ji} = w_{ji} [1 - \Psi(h - 1/2)]$, con h v.a. distribuita uniformemente in $[0, 1]$. Se $g = 50\%$ (eliminazione del 25% di tutte le connessioni) si ottiene $\alpha_c = 0.09$, mentre per $g = 100\%$ $\alpha_c < 0.06$.
- **Sen e Chakrabarti** introdussero la legge di attivazione

$$x_i(t+1) = \operatorname{sgn}\left\{\sum_{j=1..N} w_{ij} [x_j(t) + \lambda x_j(t-1)]\right\}$$

simile a quella di Caianiello. La dinamica è asincrona e la legge di apprendimento è quella canonica di Hebb. Per $\lambda > 0$ questa rete ha capacità maggiore di quella della rete di Hopfield (se $\lambda = 2$ si ha $\alpha_c = 0.25$).

- Ispirandosi alle considerazioni di Braitenberg (1984) sulla dinamica delle soglie dei neuroni, **Horn e Usher** (1989) hanno proposto la legge di attivazione

$$x_i(t+1) = F_T \left\{ \sum_{j=1..N} w_{ij} x_j(t) - s_i(t) \right\}$$

$$s_i(t+1) = a s_i(t) + b x_i(t)$$

con $Pr\{F_T(x)=y\} = (1+e^{-2xy/T})^{-1}$; $y=\pm 1$. In base a simulazioni e studi analitici si è visto che la rete possiede tre fasi, **paramagnetica**, **ferromagnetica** e **periodico-caotica**, a seconda dei valori dei parametri T , a , b .

- La legge di Hebb può essere modificata allo scopo specifico di memorizzare pattern organizzati gerarchicamente su un albero ultrametrico
- **Def.** In uno spazio metrico S un insieme di punti è strutturato ad **albero ultrametrico** se un punto a *antenato* è equidistante da due punti *discendenti* appartenenti allo stesso livello di discendenza di a .
- **Th.** Nello spazio degli stati di una rete di Hopfield, $S = \{\pm 1\}^N$, definita la metrica

$$D(\mathbf{x}, \mathbf{y}) = 1 - (1/N) \sum_{i=1..N} x_i y_i = 1 - C(\mathbf{x}, \mathbf{y}) = 2H(\mathbf{x}, \mathbf{y})/N$$

il seguente metodo permette di costruire un albero ultrametrico a due livelli con 1 antenato e R discendenti:

- Scegliere il pattern antenato \mathbf{x}^a con media nulla, $(1/N) |\sum_{i=1..N} x_i^a| < 1/\sqrt{N}$;
- Definire i pattern discendenti \mathbf{x}^s ($s=1, \dots, R$) come $x_i^s = x_i^a y_i^s$, con y_i^s v.a. avente distribuzione $Pr\{y_i^s = \pm 1\} = \frac{1}{2}(1 \pm a)$; $a < 1$, $i=1, \dots, N$;

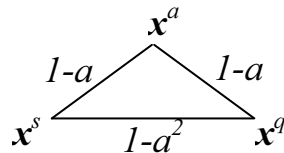
Si dimostra infatti che la distanza tra antenato e discendenti vale

$$D(\mathbf{x}^s, \mathbf{x}^a) = 1 - (1/N) \sum_{i=1..N} x_i^s x_i^a = 1 - (1/N) \sum_{i=1..N} (x_i^a)^2 y_i^s =$$

$$1 - (1/N) \sum_{i=1..N} y_i^s \approx 1 - E\{y_i^s\} = 1 - [\frac{1}{2}(1-a) - \frac{1}{2}(1-a)] = 1-a$$

mentre la distanza tra due discendenti vale, con calcoli analoghi,

$$D(\mathbf{x}^s, \mathbf{x}^q) = 1 - (1/N) \sum_{i=1..N} x_i^s x_i^q = 1-a^2$$



Per memorizzare questi $R+1$ pattern, strutturati su un albero ultrametrico a due livelli, si definisce la matrice delle connessioni nel seguente modo

$$w_{ij} = (1/N) x_i^a x_j^a [1 + (1/\delta) \sum_{s=1..R} (y_i^s - a)(y_j^s - a)]$$

con $1 - a^2 \leq \delta \leq (1 - a^2)(1 + a)/a$.

Abbiamo due casi:

- a) $\delta = 1 - a^2$. Nessuno dei pattern è uno stato di equilibrio
- b) $\delta = 1 - a^2$. Il modello si comporta come una rete di Hopfield in cui siano immagazzinati $R+1$ pattern
- c) $\delta > 1 - a^2$. $\exists \alpha_{c1}, \alpha_{c2} \mid$ se $\alpha \equiv (R+1)/N > \alpha_{c1}$ nessun pattern memorizzato è uno stato di equilibrio stabile della fase di richiamo; se $\alpha_{c2} < \alpha < \alpha_{c1}$ solo \mathbf{x}^a è punto di equilibrio stabile; se $\alpha < \alpha_{c2}$ anche $\mathbf{x}^1, \dots, \mathbf{x}^R$ diventano punti di equilibrio stabile.

E' quindi possibile fare in modo che la rete, durante la fase di richiamo, risponda solo agli stimoli sovraordinati di una gerarchia, in forte analogia con la capacità che la memoria umana ha di ricordare le proprietà generali di una classe di stimoli, dimenticando i casi particolari.

- Per costruire un albero ultrametrico a due livelli con P antenati, ciascuno con un certo numero di discendenti:
 - a) Scegliere i pattern antenati \mathbf{x}^p con media nulla ($p = 1, \dots, P$);
 - b) Definire i pattern discendenti $\mathbf{x}^{p,s}$ ($s=1, \dots, R_p$) come $x^{p,s}_i = x^p_i y^{p,s}_i$, con $y^{p,s}_i$ v.a. avente distribuzione $Pr\{y^{p,s}_i = 1\} = 1/2(1 + a_p)$; $Pr\{y^{p,s}_i = -1\} = 1/2(1 - a_p)$ ($a_p < 1$, $i=1, \dots, N$);

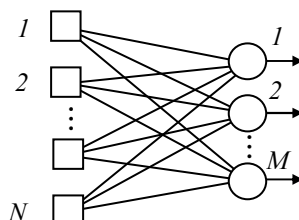
La legge di Hebb dev'essere generalizzata come segue:

$$w_{ij} = \sum_{p=1..P} (1/N) x^p_i x^p_j [1 + (1/\delta) \sum_{s=1..R_p} (y^{p,s}_i - a_p)(y^{p,s}_j - a_p)]$$

- Un limite è che δ deve essere deciso prima della fase di memorizzazione.
- Notiamo che nei casi pratici difficilmente si verificherà la condizione esatta di ultrametricità, invece avremo sovrapposizioni leggermente diverse tra discendenti dello stesso antenato e appartenenti allo stesso livello. In tal caso assegneremo al parametro a (o ai parametri a_p nei casi più complessi) la media delle sovrapposizioni. Inoltre per il calcolo dei coefficienti di connessione si può usare $y^{p,s}_i = x^p_i x^{p,s}_i$.
- Per superare la limitazione principale della rete di Hopfield possono essere usate unità del tipo $\Sigma\Pi$. **Prados** (1988) ha dimostrato che reti simili possono memorizzare fino a 2^{N-1} pattern.

ASSOCIATORI LINEARI

- Gli **associatori lineari** sono reti di unità lineari senza soglia disposti in due strati, quello di ingresso (N unità) e quello di uscita (M unità), uniti da connessioni interstrato feedforward.



- I più semplici associatori lineari sono quelli che agiscono solo su pattern **normalizzati** (rappresentati da vettori con modulo 1).
- Se desideriamo che al pattern di ingresso \mathbf{x} sia associato il pattern di uscita \mathbf{y} , diamo un incremento

$$\Delta w_{ij} = y_i x_j$$

Nel caso di una memoria **autoassociativa**, in cui ogni pattern deve richiamare se stesso, o meglio il suo prototipo non distorto, sarà $M=N$ e $\Delta w_{ij} = x_i x_j$.

- *Th.* il richiamo perfetto di un autoassociatore lineare è possibile solo se i pattern di ingresso sono tutti mutualmente ortogonali, cioè $\mathbf{x}^p \cdot \mathbf{x}^q = \delta_{pq}$.
- Possiamo scrivere sinteticamente la matrice delle connessioni al termine della fase di apprendimento (R item in tutto) definendo la matrice $N \times R$ dei pattern di ingresso $\mathbf{X} = [\mathbf{x}^1 \mathbf{x}^2 \dots \mathbf{x}^R]$ e la matrice $M \times R$ dei rispettivi pattern associati in uscita $\mathbf{Y} = [\mathbf{y}^1 \mathbf{y}^2 \dots \mathbf{y}^R]$.

Allora

$$\mathbf{W} = \mathbf{Y} \cdot \mathbf{X}^T$$

La fase di richiamo consiste nell'applicare un pattern di stimolo \mathbf{z} , ottenendo il pattern di uscita

$$\mathbf{u} = \mathbf{W}\mathbf{z}$$

- Se applichiamo in ingresso uno dei pattern memorizzati, $\mathbf{z} = \mathbf{x}^q$

$$u_i = \sum_{j=1..N} w_{ij} x_j^q = \sum_{j=1..N} \sum_{s=1..R} y_i^s x_j^s x_j^q = \sum_{s=1..R} y_i^s \mathbf{x}^s \cdot \mathbf{x}^q = \sum_{s=1..R} y_i^s \delta_{sq} = y_i^q \Leftrightarrow \mathbf{u} = \mathbf{y}^q$$

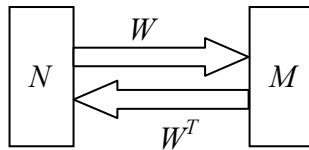
- Per superare la limitazione della necessità di pattern ortogonali, all'inizio degli anni '70 **Kohonen** propose la seguente generalizzazione

$$\mathbf{W} = \mathbf{Y} \cdot \mathbf{X}^\sim$$

Per l'autoassociatore sarà ovviamente $\mathbf{W} = \mathbf{X} \cdot \mathbf{X}^\sim$. Nel caso di pattern di ingresso ortogonali $\mathbf{X}^T \cdot \mathbf{X} = \mathbf{I}$, e la regola della pseudoinversa si riduce alla regola classica.

- Come l'associatore tradizionale, anche l'**associatore lineare di Kohonen** presenta delle limitazioni. La prima è che introducendo connessioni feedback, ovvero passando nuovamente l'uscita sull'ingresso più volte, non si ottengono miglioramenti nell'errore del pattern richiamato. Ciò è dovuto principalmente al fatto che le unità sono lineari.
- **B.Kosko** (1987) ha introdotto nonlinearità nelle funzioni di attivazione (**bidirectional associative memory, BAM**). Nel modello **BAM** la matrice delle

connessioni feedforward coincide con $W = Y \cdot X^T$, mentre la matrice delle connessioni feedback è W^T . Stavolta le unità sono binarie a soglia ($x = \pm 1$), e non lineari. Il funzionamento può essere sincrono o asincrono.

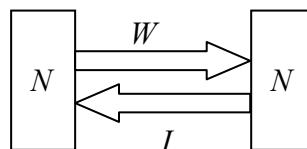


Supponendo per semplicità soglie nulle, nella fase di richiamo si applica un pattern di stimolo z (è un pattern **bipolare**, cioè di valori ± 1), ottenendo il pattern di uscita $u = \text{sgn}(Wz)$, quindi sull'ingresso ritorna il pattern $z' = \text{sgn}(W^T u)$, e sull'uscita $u' = \text{sgn}(Wz')$ ecc. finché il sistema non si stabilizza su uno stato di equilibrio, che definisce il pattern richiamato.

Il limite del modello *BAM* è la presenza di stati spuri. Il numero di associazioni memorizzabili non può superare $\min\{N, M\}$.

MEMORIZZAZIONE DI PATTERN SPAZIO-TEMPORALI

- Per la memorizzazione di pattern distribuiti spazialmente, tipicamente su una porzione di piano, la codifica in un vettore binario è abbastanza semplice. La cosa è meno semplice per la codifica di pattern temporali. In questi casi si dispone di una sequenza ordinata di pattern x^1, \dots, x^s, \dots (p.es. vettori con N componenti ± 1). La rete, stimolata con x^s , deve rispondere richiamando x^{s+1} .
- La **TAM** (**Temporal Associative Memory**) è una *BAM* capace di memorizzare sequenze periodiche (sia R il periodo). Ha unità binarie a soglia nulla, N nello strato di ingresso e N in quello di uscita.



- Essendo

$$\Delta w_{ij} = x^{s+1}_i x^s_j$$

la matrice delle connessioni feedforward è data da

$$w_{ij} = x^2_i x^1_j + x^3_i x^2_j + \dots + x^R_i x^{R-1}_j + x^1_i x^R_j \quad (i, j = 1, \dots, N)$$

mentre la matrice delle connessioni feedback è la matrice identità I . Se si utilizza una funzione di attivazione $u = \text{sgn}(Wz)$, si dimostra che per $z = x^s$ l'uscita vale

proprio $\mathbf{u} = \mathbf{x}^{s+1}$. La stessa uscita è poi riportata in ingresso, infatti $\mathbf{z}' = \text{sgn}(\mathbf{I}\mathbf{u}) = \mathbf{u}$. Poi l'ingresso produce in uscita \mathbf{x}^{s+2} , e così via fino a $\mathbf{x}^R, \mathbf{x}^1, \mathbf{x}^2 \dots$. Una proprietà interessante è che la moltiplicazione per \mathbf{W}^T produce i vettori della sequenza in ordine inverso, sicché $\mathbf{x}^{s-1} = \text{sgn}(\mathbf{W}^T \mathbf{x}^s)$. Questa proprietà si può dedurre partendo dalla $\Delta w_{ij} = x^{s-1}_i x^s_j$ (il pattern s -esimo deve produrre l' $s-1$ -esimo) e verificando che la matrice ottenuta è proprio \mathbf{W}^T .

- Oss. Si noti che questo modo di procedere è del tutto generale. Se vogliamo che il pattern di ingresso x_j ($j=1, \dots, N$) produca il pattern p_i ($i=1, \dots, M$), si porrà $\Delta w_{ij} = p_i x_j$
- Un tipo di rete di Hopfield, a dinamica sincrona, prevede una matrice di connessione

$$J_{ij} = w_{ij} + w^S_{ij} = (1/N) \sum_{s=1..R} x^s_i x^s_j + (\lambda/N) \sum_{s=1..R} x^{s+1}_i x^s_j$$

Per $\lambda > 1$ l'evoluzione libera a partire da un pattern \mathbf{x}^s passa per tutti i pattern della sequenza, a partire dall' s -esimo, ciclicamente. Cioè

$$\mathbf{x}(t) = \sum_{k=0.. \infty} \mathbf{x}^{(s+k) \bmod R} \delta(t - t_0 - k)$$

Si noti che la presentazione dei pattern dura solo un passo. Inoltre in dinamica asincrona il modello non funziona correttamente, e i pattern \mathbf{x}^s devono essere mutuamente ortogonali.

Se i pattern usati non sono ortogonali si deve modificare la formula delle matrici di connessione come segue

$$J_{ij} = (1/N) \sum_{s=1..R} x^s_i y^s_j + (\lambda/N) \sum_{s=1..R} x^{s+1}_i y^s_j, \quad \text{con } y^s_j = \sum_{r=1..R} C(\mathbf{x}^s, \mathbf{x}^r) x^r_j$$

Per far sì che la presentazione di ciascun pattern nella fase di richiamo duri τ passi, **Kleinfeld e Sompolinsky** hanno modificato l'espressione del potenziale di attivazione aggiungendo un termine detto **potenziale lento**,

$$P_i(t) \equiv V_i(t) + V^S_i(t) \equiv \sum_{j=1..N} w_{ij} x_j(t) + \sum_{j=1..N} w^S_{ij} x_j(t-\tau)$$

Una generalizzazione ulteriore di **Herz, Sulzer, Kuhn e Van Hemmen** produce una distribuzione dei ritardi temporali τ variabile da un'unità all'altra. In tal caso la rete ha bisogno di un apprendimento iniziale, in $[t_0, t_1]$. Se $\xi(t)$ è il pattern presentato al passo $t \in [t_0, t_1]$,

$$w^S_{ij}(\tau) = (\varepsilon(\tau)/N) \sum_{t=t_0..t_1} \xi_i(t) \xi_j(t-\tau)$$

una forma generalizzata della legge di Hebb (tipicamente $\varepsilon(\tau)=1$).

Nella fase di richiamo, per $t > t_1$, il potenziale lento e il potenziale di attivazione sono

$$V_i^S(t) \equiv \sum_{j=1..N} \sum_{\tau=t-1..t-1} w_{ij}^S(\tau) x_j(t-\tau)$$

$$P_i(t) \equiv (1-\gamma)[V_i(t) + V_i^S(t)] + \gamma I_i(t)$$

dove $I(t)$ è un pattern di stimolazione variabile nel tempo, e γ il parametro di sensibilità della rete ad esso.

ESPERIMENTI DI PSICOLOGIA ARTIFICIALE SULLA MEMORIA

- Tra i paradigmi sperimentali:

- 1) **esperimenti sulla memorizzazione di associazioni tra coppie di item privi di significato**. In questo caso possiamo usare gli associatori lineari, la *BAM*, la rete di Hopfield con le sue varianti. Occorre procedere con:
 - a) codifica degli item in vettori bipolari ($x_i = \pm 1$). Per assicurare media e overlap trascurabili, per la rete di Hopfield, si possono prevedere componenti aggiuntive nei vettori di codifica.
 - b) memorizzazione delle associazioni
 - c) decadimento spontaneo durante l'intervallo di ritenzione (tempo tra la fine della memorizzazione e l'inizio del test di richiamo). P.es. con

$$w_{ij}(t+1) = (1-\alpha) w_{ij}(t)$$

La **costante di decadimento** α dev'essere estrapolata dalla curva di oblio (frazione delle risposte corrette C_i vs. tempo di ritenzione t_i), notoriamente approssimata con

$$C(t) = e^{-\alpha_0 t}, \quad \alpha_0 = - (\sum_{i=1..n} t_i \ln C_i) / (\sum_{i=1..n} t_i^2)$$

Il valore di α si calcola col metodo iterativo dei minimi quadrati.

I risultati ottenuti mostrano che le NN hanno prestazioni molto più scadenti rispetto ai soggetti umani, ma anche molto insensibili all'eliminazione delle unità.

- d) presentazione di un item di richiamo.
 - e) innesco dell'evoluzione libera della rete fino al raggiungimento dell'equilibrio
 - f) decodifica del pattern di equilibrio e confronto col secondo membro della coppia di item associati; determinazione dell'errore
- 2) **esperimenti sulla memorizzazione di liste di item privi di significato**. Si utilizzano le reti capaci di memorizzare sequenze di item. Lo stimolo di richiamo che funge da innesco dev'essere quello con minore sovrapposizione

rispetto a tutti gli altri per simulare un recupero spontaneo. Le prestazioni sono sensibili al rispetto del requisito di ortogonalità.

- 3) **esperimenti sull'interferenza tra liste di item.** Le reti usate sono come al punto Occorre procedere con:
- a) memorizzazione della **lista di studio**
 - b) intervallo di latenza, con decadimento spontaneo dei coefficienti di connessione
 - c) memorizzazione della **lista test**.
 - d) nuovo intervallo di ritenzione
 - e) fase di richiamo
- 4) **esperimenti sulla memoria semantica.** Sono gli esperimenti più difficili da simulare, in quanto non si conosce a fondo questo tipo di memoria. Le reti semantiche teorizzate dagli psicologi (p.es. il modello di Collins e Loftus) possono essere implementate come NN. Esistono tuttavia due possibili interpretazioni, con vantaggi e svantaggi, su cosa debba corrispondere a un concetto, **localista** (una singola unità) e **distribuita** (un attrattore). Occorre procedere con:
- a) memorizzazione di una struttura gerarchica di concetti
 - b) intervallo di ritenzione, con decadimento spontaneo dei coefficienti di connessione
 - c) presentazione della frase che esprime una connessione tra un oggetto (pattern x^1) e la categoria di appartenenza (pattern x^2) (p.es. “il canarino è un uccello”). Si tratta di presentare nell'ordine i due pattern
 - d) innesco dell'evoluzione libera della rete fino a quando, all'istante t ,

$$C(x(t), x^2) > C(x(t), x^1)$$

Qui t viene a corrispondere al tempo di reazione.

MODELLI PSICOLOGICI DELL'APPRENDIMENTO

- Def. (**E.R.Hilgard, G.H.Bower**. *Le teorie dell'apprendimento*) “l'apprendimento è il processo con cui si origina o si modifica una attività reagendo a una situazione incontrata, ammesso che le caratteristiche del cambiamento non possano essere spiegate sulla base di tendenze a rispondere innate, di maturazione o di stati temporanei dell'organismo (p.es. fatica, droghe, ecc.).”
- Per quanto riguarda l'accezione del termine nel contesto delle NN, l'apprendimento si riferisce a reti che hanno forte interazione con gli stimoli esterni.
- Distinguiamo le **teorie stimolo-risposta** (S-R) e le **teorie cognitive**.
- La teoria stimolo-risposta di Thorndike, con la nozione di una risposta R che funge da rinforzo per l'associazione S-R, è alla base del paradigma delle NN con **apprendimento supervisionato**. Si applica alle reti in cui sia distinguibile uno

strato di ingresso e uno di uscita. In queste reti esiste un supervisore che genera un segnale di rinforzo, funzione della differenza tra la risposta emessa e la risposta corretta in corrispondenza dell'ingresso, capace di modificare i coefficienti di connessione della rete. Lo stato di soddisfazione corrisponderebbe all'inverso dell'errore quadratico globale.

La teoria di Guthrie, con le nozioni di associazione per contiguità temporale degli stimoli con le risposte, e dell'apprendimento come processo non diretto a uno scopo, è alla base del paradigma delle NN con **apprendimento non supervisionato**.

In tal caso non ci sono risposte corrette da rinforzare. La rete evolve autonomamente scoprendo correlazioni tra gli stimoli, variabili nel tempo.

La teoria di Skinner, con la nozione di apprendimento come condizionamento, tramite rinforzi, di un comportamento operante, non ha ispirato modelli di NN. Il fatto è che manca una definizione quantitativa coerente di *rinforzo*.

La teoria di Hull considera l'apprendimento come formazioni di associazioni per contiguità temporale, ma solo in presenza di rinforzo, definito come riduzione del bisogno.

- Secondo la teoria *proto-cognitiva* di Tolman apprendere non è formare associazioni S-R, ma individuare il *significato* dei *segni* presenti nell'ambiente in funzione di uno scopo. L'associazione da stabilire non è dunque tra stimoli e risposte, ma tra segni e significati.

La teoria della Gestalt ha avuto il merito di mettere in luce il ruolo dei fattori globali nei processi di apprendimento. Questo punto di vista non ha ancora trovato espressione in un modello di NN.

La teoria di Piaget vede l'apprendimento come un'interazione tra individuo e ambiente verso una condizione di equilibrio, dove il primo risulta maggiormente adattato al secondo.

In generale le teorie cognitive non hanno fornito una teoria vera e propria dell'apprendimento. Questo è visto come inseparabile da memoria, percezione, affettività, capacità sensomotoria, ecc. aspetti troppo complessi per trovare corrispondenze nei modelli di NN finora noti.

IL PERCEPTRONE

- *Def.* Il **problema della classificazione** è quello di associare un'assegnazione del vettore

$$\mathbf{x} = (x_1, \dots, x_M)^T \quad 0 \leq x_j \leq 1$$

a una delle classi C_i ($i = 1, \dots, N$).

- *Def.* Un **training set** (o **data set**, o **campione**) è un insieme di vettori $\{\mathbf{x}^1, \dots, \mathbf{x}^R\}$ per i quali già esiste l'associazione con una delle classi.
- Formalmente, il nostro algoritmo di classificazione è una funzione

$$y_i = f(\mathbf{x}, \mathbf{w}) \quad i=1, \dots, N$$

dove \mathbf{w} è un vettore di parametri, e $y_i = 1(0) \Leftrightarrow \mathbf{x}$ (non) è classificato in C_i .

- **Esempio.** Un'applicazione è il *fitting* di una curva polinomiale:

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \dots + w_r x^r = \sum_{j=0..r} w_j x^j$$

Il *data set* consiste di R punti (x^s, t^s) . Il procedimento consiste nel determinare \mathbf{w} in modo da minimizzare una qualche funzione d'errore, per esempio

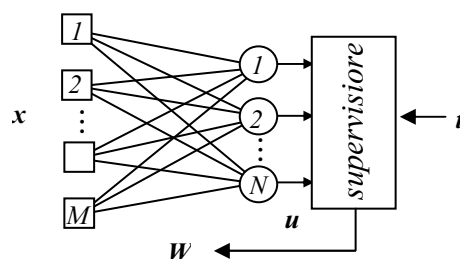
$$E(\mathbf{w}) = \frac{1}{2} \sum_{s=1..R} [y(x^s, \mathbf{w}) - t^s]^2$$

Notiamo che y è non lineare in x , ma lineare in \mathbf{w} . Funzioni che sono lineari nei parametri adattativi sono dette **modelli lineari**.

- In una rete neurale \mathbf{w} sono i coefficienti delle connessioni $\mathbf{W} = [w_{ij}]$, e la loro determinazione sulla base del *data set* è detta **addestramento** (*training*) o **apprendimento** (*learning*).
- La minimizzazione di E è un caso di *apprendimento supervisionato*, in quanto sono noti i valori *target* $\{t^s\}$. In altri problemi, come la determinazione della distribuzione $P(\mathbf{x})$, in cui non si dispone di $\{t^s\}$, si tratta di un *apprendimento non supervisionato*. Nel caso di **apprendimento rinforzato** (Hertz, 1991) non si fornisce un set di output desiderati $\{t^s\}$, ma si accetta o meno l'output prodotto spontaneamente.

IL PERCEPTRONE

- Il **Perceptrone**, con funzione di classificatore in N classi di pattern in $[0, 1]^M$, è una rete supervisionata a due strati, con M unità di ingresso, N unità binarie ($\{0, 1\}$) a soglia nello strato di uscita, e connessioni feedforward.



- Algoritmo di apprendimento:
 - 1) scelta casuale dei coefficienti w_{ij}
 - 2) scelta dell'errore ammissibile sull'insieme dei pattern di addestramento $\{\mathbf{x}^1, \dots, \mathbf{x}^R\}$ (*training set*), E_{max} .
 - 3) errore effettivo sul training set $E = 0$

- 4) indice pattern di addestramento $s = 1$
- 5) si forzano le unità di ingresso al pattern \mathbf{x}^s .
- 6) si calcolano le uscite della rete

$$u_i^s = \Psi(\sum_{j=1..M} w_{ij} x_j^s - S_i)$$

- 7) si aggiunge a E l'errore sul pattern \mathbf{x}^s ,

$$\Delta E = ||\mathbf{u}^s - \mathbf{t}^s||/R$$

- 8) si aggiunge ai coefficienti di connessione la quantità

$$\Delta w_{ij} = \eta x_j^s |u_i^s - t_i^s|$$

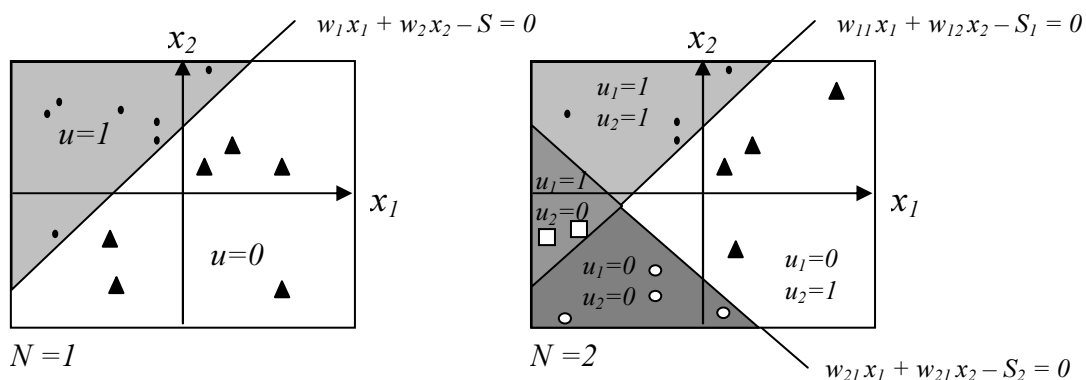
- 9) se $s < R$, $s=s+1$ e torna a 5)
- 10) se $E > E_{max}$ torna a 3)
- 11) fine

- Def. η è il **parametro di apprendimento**, positivo e di solito <1
- Def. I passi compresi tra 3) e 9) sono detti **epoca** (ciclo di presentazione di tutto il training set).
- L'algoritmo converge verso una configurazione della rete che classifica tutti i pattern del training set nei limiti dell'errore prefissato, ma solo se il problema è linearmente separabile nello spazio di ingresso. Per chiarire questo concetto consideriamo una rete con $M=2$ ingressi e $N=1$ uscita

$$u = \Psi(w_1 x_1 + w_2 x_2 - S)$$

Nel piano (x_1, x_2) l'equazione $w_1 x_1 + w_2 x_2 - S = 0$ rappresenta una retta.

L'apprendimento del perceptrone è una variazione dei parametri w_1 e w_2 che porta la retta a segregare nel modo migliore possibile ciascun punto del training set in uno dei due semipiani, secondo l'attribuzione da parte del supervisore all'una o all'altra classe. Quindi il perceptrone non potrà classificare correttamente classi di pattern che non siano linearmente separabili.



Con $N > 1$ unità di uscita aumentano le capacità di segregazione della rete, nei vari settori di piano definiti da N rette. In generale avremo N iperpiani in uno spazio M -dimensionale.

La funzione XOR (che richiede $N=1$ in quanto è una funzione a una sola uscita) non è linearmente separabile. Per ottenere la funzione XOR occorre aggiungere uno strato nascosto tra quello di ingresso e quello di uscita.

- L'algoritmo di **B. Widrow e M.E. Hoff** permette ai Perceptroni a due strati con unità di uscita di tipo qualunque ($u_i = f(P_i)$) la convergenza su una configurazione a minimo errore su un training set assegnato.
L'errore su tutto il training set è quadratico e vale

$$E = \frac{1}{2} \sum_{s=1..R} \sum_{i=1..N} (u_i^s - t_i^s)^2$$

La regola di aggiornamento dei coefficienti di connessione somma ai coefficienti correnti la quantità

$$\Delta w_{ij} = -\eta \partial E / \partial w_{ij}$$

Questa regola esprime il metodo della **discesa lungo il gradiente** (*gradient descent*). La precedente può essere esplicitata nel modo seguente

$$\Delta w_{ij} = -\eta \sum_{s=1..R} [(u_i^s - t_i^s) f'(\sum_{j=1..M} w_{ij} x_j^s - S_i) x_j^s]$$

A seconda della forma di f avremo diversi casi.

- a) **funzione di trasferimento sigmoidale.** $f(P) = 1/(1 + e^{-P})$

$$f'(P) = f(P)[1 - f(P)]$$

- b) **funzione di trasferimento lineare.** $f(P) = \alpha P$

$$f'(P) = \alpha$$

- c) **funzione di trasferimento iperbolica.** $f(P) = \tanh(P)$

$$f'(P) = 1 - [f(P)]^2$$

- **Algoritmo batch:**

- 1) scelta casuale dei coefficienti w_{ij}
- 2) scelta dell'errore ammissibile sull'insieme dei pattern di addestramento $\{\mathbf{x}^1, \dots, \mathbf{x}^R\}$, E_{max} .

- 3) errore quadratico globale $E = 0$
- 4) indice pattern di addestramento $s = 1$
- 5) si forzano le unità di ingresso al pattern \mathbf{x}^s
- 6) si calcolano le uscite della rete

$$u_i^s = f(\sum_{j=1..M} w_{ij} x_j^s - S_i)$$

- 7) si aggiunge a E l'errore sul pattern \mathbf{x}^s ,

$$\Delta E = ||\mathbf{u}^s - \mathbf{t}^s|| = \sum_{i=1..N} (u_i^s - t_i^s)^2$$

- 8) se $s < R$, $s=s+1$ e torna a 5)
- 9) se $E < E_{max}$ fine
- 10) si aggiunge ai coefficienti di connessione

$$\Delta w_{ij} = -\eta \sum_{s=1..R} [(u_i^s - t_i^s) f'(\sum_{j=1..M} w_{ij} x_j^s - S_i) x_j^s]$$

- 11) torna a 3)

- Il problema di questo algoritmo è che non garantisce la convergenza a un minimo assoluto dell'errore quadratico globale per qualsiasi assegnazione iniziale dei coefficienti \mathbf{W} . Il sistema può convergere infatti su un qualsiasi minimo locale.
- La regola di Widrow-Hoff esposta prevede il calcolo dell'errore globale su tutto il *training set* prima di ogni aggiornamento di \mathbf{W} (modalità **batch** o **cumulativa**). E' ammesso pure il caso in cui l'aggiornamento viene fatto sull'errore relativo a ciascun pattern (modalità **on-line**).
- La fase di apprendimento è seguita dalla **fase di generalizzazione**, in cui \mathbf{W} è tenuta costante e in ingresso alla rete si possono applicare nuovi pattern, anche non appartenenti al *training set*.
- E , funzione dei parametri adattivi della rete \mathbf{W} e \mathbf{S} , è l'errore quadratico globale, ma può anche essere pensata come funzione energia. E' comunque una funzione di costo che dev'essere minimizzata nello spazio dei parametri (\mathbf{W}, \mathbf{S}) durante l'apprendimento.

BACKPROPAGATION

- Passando a considerare reti con strati intermedi di cosiddette **unità nascoste** (*hidden units*) si supera il vincolo della separabilità lineare dei problemi di classificazione risolvibili dalla rete.
- L' algoritmo che generalizza la regola di Widrow-Hoff fu trovato da **Rumelhart, Hinton e Williams**, ed è noto col nome di **error backpropagation**.
- *Th. (di Kolmogorov)*. Data una funzione continua $F : [0, 1]^M \rightarrow \mathcal{R}^N$, F può essere implementata esattamente da un Perceptrone feedforward a tre strati, con M unità

nello strato di ingresso, $H = 2M+1$ unità nello strato nascosto, e N unità nello strato di uscita. Alcune delle unità devono essere però non lineari.

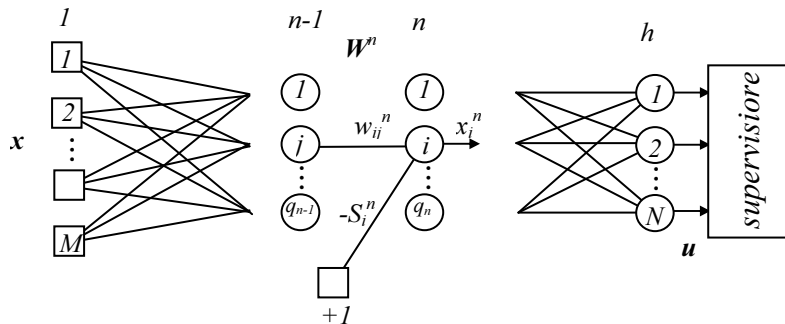
- *Th.* Un Perceptrone multistrato con sole unità lineari ($u = \alpha P \cdot S$) ha la stesse capacità di classificazione di un perceptrone con due soli strati di unità lineari.
- Siano:
 e_r^s , uscite dello strato di ingresso (durante la presentazione del pattern s -esimo)
 c_{jr} , coefficienti di connessione tra strato di ingresso e strato nascosto
 x_j^s , uscite dello strato nascosto
 w_{ij} , coefficienti di connessione tra strato nascosto e strato di uscita
 u_i^s , uscite dello strato di uscita
 t_i^s , uscita desiderata
 f , funzione di trasferimento valida per tutte le unità
 s , soglie
- Al termine della generica epoca della fase di apprendimento, anche i coefficienti c_{jr} devono essere corretti con

$$\Delta c_{jr} = -\eta \partial E / \partial c_{jr} = -\eta \sum_{s=1..R} [(x_j^s - \tilde{x}_j^s) f'(\sum_{r=1..M} c_{jr} e_r^s - s_j) e_r^s]$$

dove \tilde{x}^s è la risposta desiderata delle unità nascoste. Rumelhart, Hinton e Williams hanno dimostrato che questa quantità può essere espressa in funzione di x^s, u^s, t^s, W nel seguente modo:

$$\tilde{x}_j = \sum_{i=1..N} [(u_i^s - t_i^s) f'(\sum_{k=1..H} w_{ik} x_k^s - S_i) w_{ij}]$$

- In generale, con h strati, q_n unità nello strato n -esimo ($n=1, \dots, h$), W^n matrice $q_n \times q_{n-1}$ delle connessioni tra strato $n-1$ e n -esimo, $x^{n,s}$ vettore delle uscite dello strato n -esimo in corrispondenza dell' s -esimo pattern del *training set*, P^n vettore dei potenziali di attivazione delle unità dello strato n -esimo, S^n vettore delle soglie delle unità dello strato n -esimo,



$$\Delta w_{ij}^n = -\eta \sum_{s=1..R} \delta_i^{n,s} \cdot x_j^{n-1,s}$$

$$\delta_i^{n,s} = f'(P^{h,s}_i)(u_i^s - t_i^s)$$

se $n=h$ (strato di uscita)

$$\delta_i^{n,s} = f'(P_i^{n,s}) \sum_{r=1..q(n+1)} \delta_r^{n+1,s} \cdot w_{ri}^{n+1} \quad \text{se } n < h$$

Nella generica epoca di apprendimento, si parte dal calcolo di $\delta^{h,s}$, con una legge che è identica a quella di Widrow-Hoff. Poi si calcola ricorsivamente $\delta^{h-1,s}$, $\delta^{h-2,s}$, .. fino a $\delta^{2,s}$. Qui ci si arresta, in quanto lo strato di ingresso ($n=1$) non ha altri ingressi. Dopo aver calcolato tutti i $\delta^{n,s}$, si passa ad aggiornare i coefficienti di connessione.

- Oss. L'algoritmo non funziona usando funzioni di attivazione binarie, infatti non sono derivabili.
- Il modello ha scarsa plausibilità biologica, a causa soprattutto della necessità di un supervisore esterno.
- L'apprendimento si arresta quando il vettore gradiente, o l'errore globale, scende al di sotto di una soglia prefissata.
- Per aumentare la velocità di convergenza della rete in fase di apprendimento si può scegliere la modalità *on-line* invece che la modalità *batch*.
- Un'altra tecnica è aggiungere un **momento** nella legge di aggiornamento dei coefficienti:

$$\Delta w_{ij}^n = -\eta \sum_{s=1..R} \delta_i^{n,s} \cdot x_j^{n-1,s} + \alpha \cdot (\Delta w_{ij}^n),$$

dove $(\bullet)'$ denota la quantità \bullet relativa all'epoca precedente.

L'effetto di $\alpha \in (0,1)$ è quello di regolarizzare la convergenza della rete, evitando una dinamica di apprendimento instabile o oscillante. Se la componente $\delta_i^{n,s}$ del gradiente mantiene lo stesso segno in due iterazioni successive, l'effetto di α è quello di rinforzare la velocità di apprendimento, se invece la componente $\delta_i^{n,s}$ del gradiente cambia di segno in due iterazioni successive, l'effetto di α è quello di ridurre la velocità di apprendimento.

Tipicamente $\alpha = 0$ e η piccolo (maggiore tempo di apprendimento e minore uso di memoria), oppure α e η abbastanza grandi (p.es. $\alpha = 0.9$ e $\eta = 0.6$).

- Il metodo della backpropagation si può usare anche per modificare le soglie delle unità, $\Delta S_i = -\eta \partial E / \partial S_i$. Si arriva alla

$$\Delta S_i^n = -\eta \sum_{s=1..R} \delta_i^{n,s}$$

- Nella variante **quickpropagation** di **Fahlman** (1988) il tasso di variazione dell'errore quadratico globale rispetto ai coefficienti di connessione è ancora

$$\partial E / \partial w_{ij}^n = \sum_{s=1..R} \delta_i^{n,s} \cdot x_j^{n-1,s}$$

ma stavolta, al termine di ogni epoca, si devono memorizzare tre quantità:

$$\partial E / \partial w_{ij}^n, (\partial E / \partial w_{ij}^n)', (\Delta w_{ij}^n)'$$

Che definiscono una parabola nel piano E, w_{ij}^n , nell'ipotesi che E sia una funzione approssimativamente quadratica rispetto a ciascun coefficiente (nel caso di unità lineari lo sarebbe *esattamente*). L'idea è allora far variare w_{ij}^n in modo che vada a coincidere col vertice di questa parabola (minimo):

$$\Delta w_{ij}^n = (\Delta w_{ij}^n)' / [1 - (\partial E / \partial w_{ij}^n) / (\partial E / \partial w_{ij}^n)']$$

Nelle applicazioni pratiche occorre introdurre dei correttivi per evitare che la quantità a secondo membro diverga quando il denominatore è molto piccolo (parabola con piccolissima curvatura).

La quickpropagation si dimostra molto più veloce della backpropagation, ma è anche molto più soggetta a convergere sui minimi locali di E . per evitare quest'ultimo inconveniente si usa una funzione di trasferimento con un parametro di temperatura T , tipo $f(P) = 1/(1 + e^{-P/T})$, che per $T=0$ diventa la funzione gradino $f(P) = \Psi(P)$, mentre per $T=1$ dà la sigmoide $f(P) = 1/(1 + e^{-P})$. In tal caso si ha $f'(P) = f(P)[1 - f(P)]/T$. Tipicamente si fa variare T nel tempo secondo una determinata legge, come ad esempio

$$T(t) = T_0 - (T_0 - 1)(t - t_0)/(t_M - t_0)$$

Si tratta di un “raffreddamento”, dove t_M è il numero di passi necessario affinché T raggiunga valore 1.

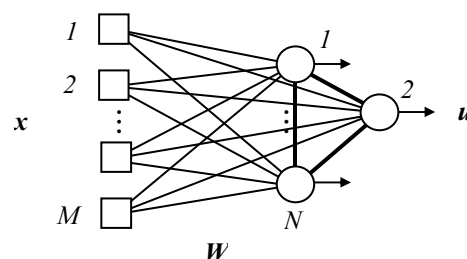
- Le **macchine di Boltzmann**, introdotte da **Sejnowski, Kienker e Hinton** (1986), hanno legge di attivazione probabilistica, con una funzione di temperatura detta **ricottura simulata** (*simulated annealing*) che permette di raggiungere sempre il minimo assoluto di E con l'ausilio di un supervisore.
- Per un apprendimento efficace occorre adottare i seguenti accorgimenti:
 - 1) Massimizzare la quantità di informazione contenuta nei pattern (usare pattern diversificati)
 - 2) Usare sempre funzioni di attivazione antisimmetriche (come la *tanh*)
 - 3) Rinormalizzare gli ingressi alla rete in modo da avere media nulla, non correlazione, covarianze simili, ampiezza minore dell'ampiezza della funzione di attivazione
 - 4) Inizializzazione casuale dei coefficienti w_{ij}^n sinaptici e delle soglie S_i^n in modo che la varianza nella scelta sia proporzionale a q_n .
 - 5) Il numero di unità nascoste non deve essere troppo piccolo (scarsa approssimazione dei dati di training) né troppo grande (peggiore capacità di generalizzazione)

RAPPRESENTAZIONE DELL'INFORMAZIONE NELLE NN

- L'analisi della distribuzione dei coefficienti di connessione di una rete con unità nascoste, al termine della fase di apprendimento, può aiutarci a capire la rappresentazione che la rete si è fatta dell'informazione contenuta nel *training set*.
- Applicati i pattern del *training set* a una rete già addestrata, l'attività di una unità nascosta potrà essere significativamente correlata con l'attività di una sola unità di ingresso (sse la rappresentazione è **locale**) o con un gruppo di unità di ingresso (rappresentazione **super-locale**), mentre l'attività di una unità di ingresso potrà essere significativamente correlata con l'attività di un gruppo di unità nascoste (rappresentazione **distribuita**) o con tutte le unità nascoste (rappresentazione **super-distribuita**).
- In presenza di rappresentazioni non locali con M elevato conviene usare la **cluster analysis** delle differenze in valore assoluto tra i coefficienti di connessione $\{w_{ij}^n\}$. Questa permette di capire qual'è l'insieme delle unità di ingresso che devono essere attivate per attivare una specifica unità nascosta.
- In generale nel Perceptrone multistrato le unità del primo strato nascosto codificano le caratteristiche *locali* dei pattern di training, mentre gli strati successivi estraggono le caratteristiche *globali*.
- Il valore medio dei coefficienti di connessione in ingresso a una unità nascosta ($\mu_i^n = \sum_{j=1..q(n-1)} w_{ij}^n / q_{n-1}$) è proporzionale al **rapporto segnale/rumore**. Maggiore è μ_i^n , più l'unità i -esima dello strato n -esimo è sensibile ai segnali in ingresso. Invece la varianza è proporzionale alla **discriminabilità**, maggiore è σ_i^n , più le differenze tra gli ingressi verranno esaltate. Questo suggerisce che la correlazione tra μ_i^n e σ_i^n possa essere usato come indice di **selettività**.

RETI AD APPRENDIMENTO COMPETITIVO

- Nell'**apprendimento competitivo** è la rete che deve catturare le regolarità statistiche della sequenza dei pattern che le gli vengono presentati. E'una modalità allineata alle teorie di Guthrie. Tra le unità si istaurano rapporti di *competizione* e di *cooperazione* che portano i parametri della rete a riprodurre la distribuzione statistica dei dati di training.



- Le reti competitive hanno connessioni feedforward interstrato ed eventualmente intrastrato (connessioni *lateral*). L'idea è aumentare i pesi delle connessioni eccitatorie verso le unità che hanno vinto la competizione. Ma chi sono i vincitori è stabilito in base a una delle seguenti strategie:

- 1) **strategia con più vincitori**. Sono attivate solo le unità di uscita con potenziale superiore alla media. Un esempio è quello di unità binarie a soglia nulla con risposta delle unità di uscita

$$u_i = \Psi[\sum_{j=1..M} w_{ij}(x_j - 1/N)]$$

- 2) strategia **WTA (winner takes all)**. Solo una delle unità di uscita emette un segnale diverso da zero, quella con potenziale massimo.
- 3) **strategia della bolla (T.Kohonen)**. Si applica alle reti dotate di localizzazione (l'indice o gli indici di una data unità ne definiscono anche la posizione in un reticolo). Le unità vincenti sono quelle che cadono in un intorno B (bolla) dell'unità con potenziale massimo, caratterizzata da un indice $i_c(\mathbf{x})$. Nelle prime versioni Kohonen distingueva una fase di *apprendimento* (su scale di tempo lunghe), con modificazione dei coefficienti di connessione, da una fase di *rilassamento* (su scale di tempo brevi) in cui l'uscita si adatta all'ingresso. E' proprio in quest'ultima fase che si formerebbe una bolla. La risposta dell' i -esima unità è

$$u_i(t + \Delta t) = u_i(t) + [I_i(t) + \sum_{j \in B} w_{ij} u_j - g(u_i)] \cdot \Delta t$$

dove $I_i(t)$ è l'ingresso (o la somma pesata degli ingressi se c'è più di uno strato), g una funzione nonlineare, mentre la sommatoria comprende il contributo delle sole unità all'interno della bolla.

Kohonen propose per i coefficienti un andamento nello spazio a cappello messicano, mutuandola dalla risposta delle cellule gangliari *centro-on* trovata da Rodieck nel sistema visivo della rana e del gatto (1965). Un esempio è

$$w_{ij} = [1 - (i-j)^2/2\sigma^2] \exp[-(i-j)^2/2\sigma^2]$$

Il **raggio della bolla** è $r = 2\sigma\sqrt{2}$.

Se si applicano pattern di ingresso con distribuzione di ampiezza con massimo in un intorno dell'indice i_c , la rete risponderà con la formazione di una bolla centrata sull'unità di uscita localizzata in posizione i_c .

P.es. applicando un ingresso binario \mathbf{x} con distribuzione

$$P_i \equiv \Pr(x_i = 1) = (1/\sigma_0 \sqrt{2\pi}) \cdot \exp[-(i-i_c)^2/2\sigma_0^2]$$

Oss. Per implementare su calcolatore la precedente distribuzione è sufficiente generare un numero con distribuzione uniforme in $[0, 1]$, se il numero generato è $< P_i$ allora $x_i = 1$, altrimenti $x_i = 0$.

La distribuzione delle attività delle unità di uscita può essere analizzata con l'ausilio del **coefficiente di appiattimento** $\gamma(t)$.

$$\gamma(t) \equiv 1/N' \sum_{i=1..N'} (u_i(t) - \mu)^4 / [1/N' \sum_{i=1..N'} (u_i(t) - \mu)^2]^2 - 3$$

dove N' è il numero delle unità con uscite superiori all'attività media, μ il valore medio delle loro uscite. Se $\gamma = 0$ la distribuzione è una Gaussiana, se $\gamma < 0$ è più appiattita, se $\gamma > 0$ è positivo è più appuntita della Gaussiana (formazione della bolla).

Per le reti con solo uno strato di ingresso e uno strato di uscita Kohonen propose l'algoritmo di **short-cut** per semplificare le cose. Non esistono più connessioni intrastrato, mentre viene aggiunta un'equazione indipendente che definisce il raggio della bolla, $r(t)$. Se il vettore dei coefficienti di connessione che fanno capo a ciascuna unità di uscita sono normalizzati ($\sum_{j=1..M} w_{ij}^2 = 1$) allora il centro della bolla corrisponde alla coordinata spaziale dell'unità di uscita con attività massima, i_c (se la condizione di normalizzazione non è soddisfatta, occorre eseguire la normalizzazione). Stabilito il centro della bolla, le unità vincitrici sono quelle di indice i tale che $|i - i_c| \leq r(t)$.

- Algoritmi di apprendimento (o **regole di auto-organizzazione**).

1) L'**algoritmo di Rumelhart-Zipser** (1985) è un caso particolare di quello più generale di **Grossberg** (1976). La strategia di competizione è la **WTA**. Le unità sono tutte binarie (0,1).

Applicato il pattern \mathbf{x}^s in ingresso, non necessariamente bipolare, gli unici coefficienti che devono essere aggiornati sono quelli che fanno capo all'unità vincente (sia la i -esima):

$$\Delta w_{ij} = g(c_i^s/n_s) - g w_{ij}$$

dove n_s è il numero di componenti non nulle di \mathbf{x}^s , $c_i^s = \Psi(x_i^s)$, g un parametro di apprendimento. A ogni passo è anche richiesto che sia $\sum_{j=1..M} w_{ij} = 1$ quindi ogni volta occorre rinormalizzare i coefficienti nel seguente modo

$$w_{ij} \rightarrow w_{ij} / (\sum_{j=1..M} w_{ij})$$

Una variante è l'**algoritmo leaky**, in cui vengono modificati tutti i coefficienti:

$$\begin{aligned} \Delta w_{ij} &= g_w (c_i^s/n_s) - g_w w_{ij} && \text{se l'unità } i \text{ vince} \\ \Delta w_{ij} &= g_l (c_i^s/n_s) - g_l w_{ij} && \text{se l'unità } i \text{ perde} \end{aligned}$$

con $g_l \ll g_w$.

Questo modello ha il difetto di avere una codifica instabile quando si applichi ripetutamente un'opportuna sequenza di pattern da classificare (non è sempre la stessa unità di uscita a vincere in corrispondenza dello stesso pattern di

ingresso). Ciò avviene in quanto la codifica è troppo sensibile alla storia della presentazione dei vari pattern.

- 2) L'**algoritmo di Nakano-Niizuma-Omori** (1989) prevede pattern di ingresso con componenti $x_i^s \in [-1, 1]$. La strategia di competizione è la *WTA*. Lo strato di ingresso non fa altro che ritrasmettere il pattern immutato. L'uscita i -esima ha potenziale

$$P_i = \sum_{j=1..M} w_{ij} x_j^s - S_i$$

Rispetto al caso precedente anche le soglie vengono modificate dall'apprendimento:

$$\begin{aligned} \Delta w_{ij} &= \alpha (x_i^s - w_{ij}) && \text{se l'unità } i \text{ vince} \\ \Delta w_{ij} &= 0 && \text{se l'unità } i \text{ perde} \end{aligned}$$

$$\begin{aligned} \Delta S_i &= \beta + \delta [M - \sum_{j=1..M} w_{ij} x_j^s]^{1/2} && \text{se l'unità } i \text{ vince} \\ \Delta S_i &= S_i (\gamma - 1) && \text{se l'unità } i \text{ perde} \end{aligned}$$

con $0 < \alpha < 1$, $0 < \gamma < 1$, $\beta > 0$, $\delta > 0$.

Con ciò anche se un'unità perde aumenta la sua probabilità di vincere in futuro, dato che la sua soglia si abbassa.

- 3) L'**algoritmo di Kohonen** (1989) nella versione *short-cut* prevede l'aggiornamento dei coefficienti di connessione relativi alle sole unità appartenenti alla bolla che esiste all'istante t , secondo la legge

$$\begin{aligned} w_{ij}(t+1) &= w_{ij}(t) && \text{se } i \notin B \text{ (i.e. } |i - i_c(t)| > r(t)) \\ w_{ij}(t+1) &= w_{ij}(t) + \alpha(t)(x_i^s - w_{ij}) && \text{se } i \in B \end{aligned}$$

generalmente il parametro di apprendimento α e il raggio della bolla r vengono fatti diminuire nel tempo. In particolare per r possiamo avere

$$\begin{aligned} r(t) &= r_i(r_f/r_i)^{t/t_m}, && (p.es. r_i=4; r_f=0.5, t_m=2000) \\ r(t) &= A - Bt, && (p.es. A=4; B=0.002) \\ r(t) &= A/(1+Bt), && (p.es. A=4; B=0.003) \end{aligned}$$

Con ciò accade che i cambiamenti iniziali sono più importanti, e interessano un intorno più ampio dell'unità vincente.

Questo algoritmo può produrre **mappe semantiche**, in modo tale che operando una partizione dei pattern di ingresso in un sotto-vettore "nome" (K componenti) e in un sotto-vettore "concetto" (M-K componenti), ad apprendimento avvenuto specifiche unità si attiveranno alla presentazione del solo nome. Si è cioè creata un'associazione tra nomi e unità. La cosa

interessante è che unità vicine corrispondono a nomi di entità con caratteristiche vicine.

- **Esempio di Algoritmo:**

In questo caso l'aggiornamento dei pesi è fatto su tutte le unità dello strato di uscita, e il raggio della bolla $r(t)$ è rappresentato dalla funzione $\sigma(n)$ nell'equazione di apprendimento. Sia il raggio della bolla che il coefficiente di apprendimento α sono aggiornati solamente ad ogni epoca di apprendimento.

- 1) scelta casuale dei coefficienti w_{ij} (tip. $\ll x_j^s$)
- 2) indice epoca $n = 0$
- 3) calcola parametro di apprendimento:

$$\alpha(n) = \alpha_0 \exp(-n/\tau_2) \quad \alpha_0 = 0.1, \tau_2 = 1000$$
- 4) calcola raggio della bolla:

$$\sigma(n) = \sigma_0 \exp(-n/\tau_1) \quad \sigma_0 = \text{raggio della rete}, \tau_1 = \tau_2 / \ln \sigma_0$$
- 5) $s = 1$
- 6) fase di competizione: l'indice dell'unità vincente è quella cui fanno capo pesi sinaptici più simili al pattern di ingresso,

$$i_c(\mathbf{x}^s) = \arg \min_{i=1..N} \|\mathbf{x}^s - \mathbf{w}_i\|$$
- 7) fase di apprendimento:

$$h_i(n) = \exp[-(i-i_c)^2 / 2\sigma^2(n)]$$

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \alpha(n)h_i(n)(\mathbf{x}^s - \mathbf{w}_i) \quad i = 1, \dots, N$$
- 8) se $s < R$, $s = s+1$ e torna a 6)
- 9) se $n < N_{epoche}$, $n = n+1$ e torna a 3)
- 10) fine

- al termine dell'apprendimento (tipicamente ~1000 epoche) la distribuzione dei pesi sinaptici rispecchia la distribuzione statistica dei pattern $\{\mathbf{x}^1, \dots, \mathbf{x}^R\}$.
- Un problema di molte reti ad apprendimento competitivo, ma anche delle reti ad apprendimento supervisionato, è la eccessiva plasticità delle connessioni, che dà luogo a instabilità di codifica. Nelle reti supervisionate a backpropagation, per esempio, una nuova sessione di apprendimento, basata su un nuovo *training set*, cancella le informazioni immagazzinate in precedenza. Non esiste una MLT.
- Con la **teoria della risonanza adattiva** (*Adaptive Resonance Theory, ART*) **Grossberg**, a partire dal 1969, ha cercato di costruire modelli con un compromesso tra plasticità e stabilità, proprio come avviene negli esseri viventi. Esistono due processi continuamente operanti, uno **bottom-up** (propagazione del pattern dall'ingresso all'uscita) e uno **top-down** (ritorno del pattern di uscita sull'ingresso, e confronto col pattern di ingresso). Questi algoritmi, molto complessi, sono stati usati per costruire modelli della percezione visiva, della memorizzazione di liste, del movimento intenzionale, per il riconoscimento radar, ecc.

- Per ovviare alle complicazioni dei modelli ART, R.Hecht-Nielsen ha ideato le **counterpropagation networks**. La versione più semplice ha 3 strati, uno di ingresso (M unità), uno intermedio (H unità) e uno di uscita (N unità).

L'apprendimento consiste nelle seguenti fasi

- 1) *fase di Kohonen*. Per ogni pattern del *training set* presentato in ingresso si determina l'indice dell'unità intermedia con minimo valore $w_{ij} \cdot x_j$. Solo l'uscita di questa unità ("vincitrice") viene posta a 1, le altre a 0. Le connessioni facenti capo a questa unità vengono modificate secondo la:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha(t)[x_j(t) - w_{ij}(t)]$$

con $\alpha(t)$ decrescente nel tempo.

Questa fase ha termine solo quando tutti i coefficienti non cambiano più.

- 2) *Fase di Grossberg*. Vengono nuovamente presentati i pattern del training set, e l'uscita è calcolata

- Le reti con leggi di apprendimento competitivo si prestano meglio delle reti basate su backpropagation a rappresentare i processi cognitivi, in particolare quelli di categorizzazione. Tuttavia è necessario dotarle di un giusto compromesso tra plasticità e stabilità.

METODI DI VALUTAZIONE DELLE PERFORMANCE

- Un metodo di valutazione delle performance, utile per definire i valori ottimali dei parametri di un modello, o per stabilire le capacità di classificazione di una rete, è quello basato sulle curve **ROC (Receiver Operating Characteristic)**. Per una rete con una sola unità binaria in uscita (una sola classe di appartenenza C), sottoposta a una fase di apprendimento e a una successiva fase di generalizzazione sui pattern \mathbf{x}^g definiamo:

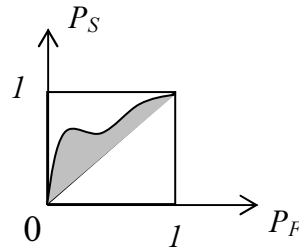
$$P_S = S/(S+M) \quad \text{probabilità di successo (o True Positive Ratio)}$$

$$P_F = F/(F+E) \quad \text{probabilità di falso allarme (o False Positive Ratio)}$$

dove S è il numero di **successi** ($\mathbf{x}^g \in C, y=1$), F numero dei **falsi allarmi** ($\mathbf{x}^g \notin C, y=1$), M numero dei **colpi mancati** ($\mathbf{x}^g \in C, y=0$), E numero delle **eliminazioni corrette** ($\mathbf{x}^g \notin C, y=0$). Notiamo che $0 < P_S, P_F < 1$.

La **curva ROC** è il grafico dei punti (P_S, P_F) al variare di un parametro della rete (p.es. soglie, numero di unità nascoste, ecc.). Se la curva ROC cade sulla bisettrice del 1 quadrante la rete non ha alcuna capacità di categorizzazione. Nei casi reali la curva è compresa nel mezzo quadrato di vertici $(0,0)$, $(1,0)$, $(1,1)$, è il suo integrale definito su $(0,1)$ fornisce una misura delle capacità di categorizzazione.

Se $N > 1$ si dovranno ricavare N curve ROC e poi calcolare la media aritmetica degli integrali.



RETI DI NEURONI ANALOGICI

- I biologi preferiscono descrivere l'attività u in termini di *frequenza media di emissione di impulsi* del singolo neurone, f , e il potenziale P è identificato col potenziale di membrana, V . Si noti che $f(t)$ e $V(t)$ sono variabili continue a tempo discreto. La frequenza di scarica (misurata in s^{-1} e compresa in $[0, 1]$) è legata al potenziale dalla relazione

$$f = g(V)$$

essendo g il **guadagno**, monotonamente crescente in tra 0 e 1, p.es. una sigmoide

$$f(V) = 1/(1 + e^{-bV}).$$

- Il potenziale V varia a causa dei seguenti fattori
 - 1) Correnti indotte dall'attività dei neuroni presinaptici
 - 2) Dispersione della corrente attraverso la membrana, a resistività non nulla, R
 - 3) Correnti iniettate dall'esterno, I

Se il rilascio di neurotrasmettitore è proporzionale a f e si trascurano gli effetti modulatori dei potenziali postsinaptici dovuti p.es. a ritardi sinaptici e fattori di forma del neurone, un'equazione dinamica può essere (Grossberg, 1969; Grossberg e Pepe, 1971; Hopfield, 1984; Sejnowski, 1981)

$$C_i dV_i/dt = \sum_{j \neq i} w_{ij} g(V_j) - V_i/R_i + I_i$$

- Questa equazione è deterministica, perché si è considerata la *frequenza media* di scarica. In realtà l'attività dei neuroni è un processo stocastico (la probabilità di scarica di un impulso in un tempo dt sarà $g(V)dt$). Il comportamento collettivo di un modello analogico siffatto è qualitativamente simile al modello binario asincrono (Hopfield, 1984).

ANALISI DELLA FORMA

- Def. Un **parametro di forma** (*shape parameter*) è una quantità che non dipende da trasformazioni geometriche quali traslazione, rotazione, cambiamento di dimensioni, riflessione.
- Consideriamo oggetti 2D, oppure immagini 2D, proiezioni o sezioni di oggetti 3D
- Assumiamo che questi oggetti siano insiemi (nel senso della teoria matematica degli insiemi) chiusi e compatti (in senso topologico), cioè **figure** (*figures*).
- Tre sono i metodi seguiti nell'analisi delle forme:
 - a) approccio funzionale - esprime i contorni delle figure come funzioni
 - b) approccio basato sulla teoria degli insiemi – p.es. la teoria dei frattali
 - c) descrizione puntuale - sostituisce alla figura un insieme di punti, solitamente giacenti sul bordo, in base a qualche proprietà fisico/geometrica, e studia la loro distribuzione con metodi statistici.
- I vantaggi dell'approccio funzionale sono:
 - 1) compressione efficiente dei dati: di solito pochi coefficienti relativi alle funzioni approssimanti bastano a una descrizione precisa della forma
 - 2) si riesce a descrivere forme anche complesse
 - 3) caratterizzazione intuitiva di molte proprietà delle forme
- Gli svantaggi:
 - 1) necessità di fissare un punto di origine, spesso arbitrario
 - 2) formule complicate per esprimere forme anche semplici
- Se la figura possiede un asse di simmetria, si può usare la distanza euclidea dei punti del contorno dall'asse di simmetria, in funzione della posizione su di esso.

.....

Appunti integrati su
MIND, MATTER, AND QUANTUM MECHANICS
Henry P. Stapp

- La logica e la storia sembrano decretare l'impossibilità di comprendere razionalmente la possibilità di interazione tra due cose così diverse come la materia, di cui è possibile un'esperienza condivisa, e il pensiero, che appartiene solamente a noi stessi.
- Materia e mente dovrebbero essere visti come manifestazioni di un'unica realtà primaria. Ma quale sarebbe la natura di questa realtà?
- Si tratta di spiegare come il contenuto di ciascun pensiero umano cosciente, così come descritto in termini psicologici, sia in relazione con i corrispondenti processi che si verificano nel cervello umano, così come descritti nel linguaggio della scienza fisica contemporanea. La conclusione di Stapp differisce da quella della maggior parte dei fisici quantistici che hanno scritto sull'argomento. Stapp chiama il suo modello esplicativo il *modello Heisenberg/James*, perché unifica l'idea di materia proposta da Werner Heisenberg con il concetto di mente di William James.
- Generalmente si è assunto che l'idea di materia usata nella Meccanica Newtoniana potesse essere applicata al funzionamento interno di un cervello. Tuttavia non sembra lecito passare dalla dinamica dei corpi celesti e della caduta dei gravi ai complessi processi chimici che si verificano nel tessuto nervoso. Anzi, l'idea classica di materia è logicamente incompatibile con la natura di molti processi essenziali al funzionamento del cervello.
- La fisica classica (**CP**) non ha un posto naturale per la coscienza: l'universo fisico è descritto in termini di:
 - a) proprietà semplici assegnate a punti dello spazio-tempo, cioè particelle e campi locali (*ipotesi locale-riduzionistica*). In particolare una particella è un oggetto localizzato nello spazio, caratterizzato da impulso ed energia. Un'onda, diffusa nello spazio, è caratterizzata da ampiezza e frequenza.
 - b) la disposizione di queste particelle e campi nello spazio ad un tempo iniziale t_0
 - c) le leggi deterministiche del moto cui campi e particelle obbediscono per $t > t_0$ (*ipotesi deterministica*). La probabilità non ha esistenza oggettiva, ma è solo una creazione della mente

Il fatto che esistano esperienze soggettive coscienti è influente sull'evoluzione dell'universo secondo la CP.

- Ma la CP ha una validità limitata, anche nello spiegare proprietà di base dei sistemi materiali, anche inorganici o inconsci.
- La teoria quantistica (QT) è di solito considerata una teoria sui fenomeni atomici. Ma la forma peculiare degli effetti quantistici implica che le idee classiche ordinarie sulla natura del mondo fisico sono profondamente sbagliate sotto aspetti che vanno molto al di là delle proprietà di singoli atomi. Inoltre, il modello di realtà fisica più accreditato oggi tra i fisici, diciamo quello di Heisenberg, implica molti effetti non classici su scala macroscopica.
- Teorie concorrenti sulla connessione mente-cervello sembrano sempre operare un qualche salto logico, una sorta di “*miracolo*”. Invece, il modello derivante dal concetto di materia proposto da Heisenberg non comporta miracoli o proprietà più speciali di quelle inerenti il modello stesso di realtà fisica di Heisenberg. La teoria individua il punto in cui nell'elaborazione nervosa del cervello si inserisce la coscienza, e spiega sia il contenuto dei pensieri coscienti che la loro efficacia causale sul movimento.
- Il modello di Stapp non è uno sviluppo teorico isolato, ma è il risultato logico di un processo storico che ha occupato gran parte di questo secolo, passando, in psicologia, dal movimento comportamentista, alla rivoluzione cognitivista, al materialismo, e in fisica dalla teoria della relatività speciale di Einstein, alla QT, al ***paradosso di Einstein-Podolsky-Rosen (EPR)***, all'elaborazione di alcuni modelli di realtà fisica che rispondono alle esigenze imposte dalla natura dei fenomeni quantistici, tra i quali quello di Heisenberg.
- Il contesto storico e logico per questi sviluppi è la delucidazione fornita da William James, alla fine del 1800, del conflitto tra la fenomenologia della mente e i precetti della fisica classica.
- Il primo punto critico è che le peculiarità della natura rivelate dai fenomeni quantistici non possono essere ignorate come effetti esoterici che appaiono solo su scala atomica. Il ***paradosso di Einstein-Podolsky-Rosen (EPR)*** è un enigma che può essere interamente espresso in termini di comportamenti di oggetti direttamente osservabili a occhio nudo.
- Nel 1935 ***Albert Einstein, Boris Podolsky e Nathan Rosen*** scrissero un celebre articolo (*Can Quantum Mechanical Description of Physical Reality Be Considered Complete?*) che condusse a ciò che oggi è considerato una proprietà inaspettata della natura: l'apparente necessità, a un qualche livello profondo, di forti azioni istantanee su grandi distanze. Questa conclusione, che è diametralmente opposta alle idee di Einstein, si deduce dalle predizioni che la QT fa in certi speciali tipi di esperimenti. Tipicamente, nello stesso tempo, ma in luoghi distanti (diciamo A e B), si effettuano due misurazioni indipendenti su un singolo sistema esteso (una *coppia* di misure). In questi casi esistono 4 coppie di misurazioni possibili, per ciascuna delle quali la QT fa delle predizioni sulla

connessione tra i risultati delle due misure. Einstein mostrò che queste asserzioni, prese insieme, entrano in conflitto con le idee ordinarie sulla natura della realtà fisica. Negli anni sono state costruite importanti generalizzazioni del paradosso EPR, e il conflitto si è approfondito. La versione più recente del paradosso EPR si basa su un esperimento di **Lucien Hardy**. La misura che ciascuno dei due sperimentatori eseguirà nel suo laboratorio sarà da lui scelta tra due misure alternative $\{U, V\}$, e il risultato di una misura sarà uno solo di due possibili risultati alternativi

$$U \rightarrow \{u_1, u_2\}; V \rightarrow \{v_1, v_2\}$$

La QT stabilisce 4 proposizioni, o predizioni:

- P1) Se in A si misurasse U e il risultato fosse u_1 , allora se in B si misurasse V , il risultato sarebbe v_1
- P2) Se in A si misurasse V e il risultato fosse v_1 , allora se in B si misurasse V , il risultato sarebbe v_2
- P3) Se in A si misurasse V e il risultato fosse v_2 , allora se in B si misurasse U , il risultato sarebbe u_2
- P4) Se in entrambe i luoghi si misurasse la U , dopo un gran numero di prove entrambe i risultati sarebbero u_1 in $1/16$ dei casi

Th. Queste quattro proposizioni, considerate nell'insieme, portano a una contraddizione se si ammette l'**assunzione di località**, secondo la quale *“la scelta finale fatta dallo sperimentatore in A su quale delle due misure effettuare non può influenzare il risultato che l'altro sperimentatore otterrà, nello stesso istante, in B”*.

Dim. See pag.7-8

- Le predizioni della QT ortodossa sono, in linea di principio, sempre asserzioni su oggetti macroscopici osservabili.
- Paradossi del tipo EPR, Scoperti negli anni '30 da Schrödinger e Einstein, non sono infrequenti in QT.
- William James definisce la psicologia come la scienza della vita mentale, comprendente *“sentimenti, desideri, conoscenze, ragionamenti, decisioni, e la volontà”*. Egli distingue due modi di unificazione: lo *spiritualismo*, che ricerca un principio semplice unificante *dietro* ai fatti psichici, e l'*associazionismo*, che ricerca elementi comuni *all'interno* dei fatti psichici.
“Il concetto di azione riflessa è certamente una delle maggiori conquiste della teoria psicologica; perché non essere radicali con essa? Perché non ammettere che, proprio come il midollo spinale è una macchina con pochi riflessi, anche gli emisferi sono una macchina con molti, e che questa è tutta la differenza? Il principio di continuità dovrebbe spingerci ad accettare questa conclusione”

James ritiene che:

- a) la coscienza sia principalmente un “***selecting agency***”, essendo sempre presente quando occorre operare una scelta tra diverse azioni
 - b) ogni pensiero, per quanto complicato e strutturato, è sempre uno stato indiviso di coscienza. L’unità della coscienza è incompatibile con l’ipotesi locale-riduzionistica della CP. Prevedette importanti mutamenti in fisica.
 - c) l’introspezione, pur “*difficile e fallibile*”, è un possibile strumento di indagine
 - d) Fu uno dei tre principali sostenitori della *filosofia pragmatica*: la misura della verità di un’idea è quanto essa è funzionale alla vita. Una posizione analoga alla *interpretazione di Copenhagen* della QT.
 - e) L’universo è fatto di un’unica sostanza primordiale, l’***esperienza pura*** (*pure experience*). Secondo la QT è lo *stato evolvente dell’universo* ma con due modi di evoluzione dinamica: uno regolare e uno discontinuo. Il primo è l’evoluzione, localmente deterministica, delle potenzialità e probabilità delle cose reali
- Nel 1905 Einstein espose la ***teoria della relatività speciale***. Questa teoria modificò l’idea di natura che aveva dominato nella scienza per tre secoli, in particolare:
 - 1) l’idea newtoniana di uno spazio e un tempo assoluti e omogenei, implicante l’esistenza di un “ora” universale, definito senza ambiguità in ogni punto dello spazio
 - 2) l’idea di azione istantanea a distanza
 - Nel 1915 Einstein pubblicò la ***teoria della relatività generale***, un’impresa molto superiore alla relatività speciale. La teoria riuniva lo spazio e il tempo assoluti in uno ***spaziotempo*** assoluto. Non esistono osservatori o misure, ma un unico plenum di punti spaziotempo inosservabili collegati assieme da equazioni differenziali. Nei 10 anni di lavoro Einstein non fu mosso dall’esigenza di codificare dati, ma da un astratto “*principio di equivalenza*”, di coerenza razionale. Einstein inviò il lavoro finito a Max Born dicendo che una volta capita la teoria, non ci sarebbe stato bisogno di conferme sperimentali. In seguito furono eseguiti dei test e le predizioni della teoria furono confermate.
 - Nel 1926 Heisenberg descrisse la QT in un simposio a Berlino, al quale partecipò Einstein, il quale restò scettico.
 - La ***QT*** è una teoria statistica. La posizione e velocità di una particella classica è definita assegnando una probabilità in ogni punto dello spazio di stato, a 6 dimensioni, della particella. Questa distribuzione di probabilità varia nel tempo. Un dispositivo di rivelazione della particella può essere descritto specificando in ogni punto la probabilità che avrebbe il dispositivo di rivelare la presenza della particella se questa si trovasse esattamente in quel punto. Combinando le due distribuzioni si ottiene la probabilità per unità di tempo di un evento di rivelazione. Nel caso quantistico si ha che:

- a) L'evoluzione temporale della distribuzione di probabilità è governata da una diversa equazione del moto
 - b) La "probabilità" assegnata a ciascun punto dello spazio di stato può essere negativa
- Il punto b) fa capire che il modo di combinare le due distribuzioni è anche diverso, pur essendo il risultato lo stesso. Questo risultato è ciò che conta, non quello che sta dietro i fenomeni osservabili. Questo punto di vista pragmatico, sviluppato da Bohr, Heisenberg, Pauli e altri alla fine degli anni '20 sul significato e l'uso della QT, è detto **Interpretazione di Copenhagen**:

- a) "The task of sciences is to extend the range of our experience and reduce it to order..." (Bohr). Si noti che l'obiettivo è l'estensione della "nostra esperienza", e non la conoscenza della struttura della realtà oggettiva.
 "...un fenomeno fisico non e' tale finche' non viene osservato";
 "E' sbagliato pensare che il compito della fisica sia dire come la Natura e'.
 La fisica verte su cio' che della natura possiamo dire."
- b) L'interpretazione fisica appropriata del formalismo simbolico della QT consiste solo nelle predizioni statistiche riguardo a singoli fenomeni che si verificano sotto opportune condizioni fisiche
- c) Un sistema può essere investigato in modi complementari diversi che rivelano proprietà che non possono essere rappresentate come possedute simultaneamente dal sistema.

Improntata alla filosofia positivista, sembrò soddisfare la maggior parte dei fisici quantistici durante la prima metà del secolo, tanto da diventare negli anni '30 la QT ortodossa (in realtà gli scritti di Heisenberg e Bohr non chiariscono perfettamente e univocamente cosa sia l'interpretazione di Copenhagen).

La procedura ortodossa della QT si articola nei seguenti passi:

- 1) le specifiche A di preparazione del sistema fisico sono tradotte nella funzione d'onda $\psi_A(\mathbf{x})$ (\mathbf{x} è l'insieme delle variabili del sistema, o i suoi *gradi di libertà*)
- 2) le specifiche B di misurazione del sistema e dei possibili risultati della misura sono tradotte nella funzione d'onda $\psi_B(\mathbf{y})$ (\mathbf{y} è l'insieme delle variabili del sistema misurato)
- 3) si costruisce la funzione $U(\mathbf{x};\mathbf{y})$
- 4) si calcola l'"ampiezza di transizione"

$$\langle A|B \rangle = \int \psi_A(\mathbf{y}) U(\mathbf{x};\mathbf{y}) \psi_B^*(\mathbf{y}) d\mathbf{x}d\mathbf{y}$$

- 5) la probabilità che il risultato della misura sia quello specificato in B vale

$$P(A,B) = |\langle A|B \rangle|^2$$

Questa probabilità è proprio il limite della frequenza relativa di occorrenza del risultato previsto, al tendere a ∞ del numero di esperimenti (*interpretazione frequentista*)

- Il modo in cui le specifiche di preparazione e di misura del sistema sono tradotte nella funzioni d'onda $\psi_A(\mathbf{x})$ e $\psi_B(\mathbf{x})$ è il **problema della misura** nella QT, che fu studiato da Von Neumann.

- Ma negli anni '50 furono avanzati tre possibili modelli di cosa si verifica realmente nella natura.

- Nel 1952 **David Bohm** propose un modello del mondo fisico che, in modo essenzialmente meccanico, può riprodurre tutte le predizioni della **QFT (Teoria dei Campi Quantistici)** e spiegare il paradosso EPR:

- a) la distribuzione di probabilità esiste realmente in natura, oltre che nella mente degli scienziati
- b) esiste un insieme di mondi classici di particelle puntiformi (e/o di campi classici), cioè dello stesso tipo postulato dalla fisica classica, distribuiti in posizione e velocità secondo un'unica distribuzione di probabilità
- c) essi continuerebbero indefinitamente ad essere conformi alla stessa distribuzione, che evolve istante per istante, essendo soggetti a una **forza istantanea** (*forza di Bohm*) dipendente dalla distribuzione di probabilità assegnata
- d) la distribuzione di probabilità dinamicamente si decompone, al livello delle variabili macroscopiche, durante certi processi fisici "*di misura*" ("*measurement-type*"), in direzioni distinguibili, una sola delle quali è scelta da un meccanismo non locale. Solo tale direzione è vissuta nell'esperienza cosciente dell'uomo.

- Un secondo modello prevede che ciascuno dei risultati possibili di un processo di misura dia luogo a un'esperienza cosciente. Ciò avverrebbe in seguito alla suddivisione dell'universo in molti universi paralleli non comunicanti.
- Il terzo modello è quello rielaborato da Stapp sulla base di un'idea di Heisenberg. Il modello mantiene l'idea che la distribuzione di probabilità che si ha in QT esista realmente in natura. Questa distribuzione, e le sue rapide variazioni, danno una rappresentazione completa della realtà. Il modello fa a meno del mondo fisico classico di Bohm. Il dualismo si sposta invece sul piano della dinamica, la quale procederebbe oscillando tra due fasi

- a) evoluzione graduale secondo leggi deterministiche analoghe a quelle della fisica classica (sostanzialmente è l'equazione di Schrödinger)

- b) rapidi e incontrollati salti quantistici. Questi *eventi* selezionano una delle due alternative macroscopiche ed eliminano l'altra, e sono ciò che avviene nella realtà
 - c) L'evento può verificarsi solo dopo che la fase a) ha costruito le possibilità macroscopiche alternative tra cui dev'essere fatta la scelta
 - d) La tendenza che ciascuna possibilità ha a realizzarsi è proporzionale alla distribuzione di probabilità, realmente esistente.
- Il modello di Heisenberg è più semplice del modello di Bohm, infatti quest'ultimo deve ammettere l'esistenza di un mondo classico sottoposto all'azione di una forza non locale, e inoltre i rami della distribuzione di probabilità non scelti in qualche modo intricato continuerebbero a sopravvivere.
 - Il modello mente/cervello è basato sull'analogia tra cervelli e dispositivi di misura.
 - Il postulato principale del modello H/J è che ogni *evento cosciente* è il correlato psicologico di in uno speciale tipo di *evento di Heisenberg* nel cervello, un pattern di attività neuronale che deve essere dapprima costruito, in sovrapposizione a molti altri, da un'attività cerebrale inconscia meccanica, quindi attualizzato mediante un processo di selezione. L'evento fisico e l'evento psichico devono essere immagini l'uno dell'altro sotto un isomorfismo matematico ben preciso, che mappa eventi coscienti in una descrizione psicologica in eventi di Heisenberg in una descrizione fisica del cervello.
 - Il filosofo della mente **Daniel C.Dennett**, nel suo libro ***Consciousness Explained*** dice: "L'idea di mente come distinta... dal cervello, composta non di ordinaria materia ma di qualche altro tipo di sostanza, è *dualismo*, ed è oggi meritatamente caduto in discredito... Dai tempi del classico attacco di Gilbert Ryle a quello che lui definì il "*dogma del fantasma nella macchina*" di Cartesio, i dualisti sono stati sulla difensiva. La verità dominante, diversamente espressa e discussa, è il *materialismo*: c'è un tipo di sostanza, la *materia* (la sostanza della fisica, della chimica e della fisiologia) e la mente non è altro che un fenomeno fisico. In breve, la mente è il cervello"
 - Ryle è anche citato nel libro di **Bernard Baars**, ***A Cognitive Theory of Consciousness***: "... il filosofo Gilbert Ryle oppose argomenti assai convincenti alle entità mentali, che egli ridicolizzò come *fantasmi nella macchina* e *homunculi*. Ryle credeva che tutte le inferenze mentalistiche implicassero una mescolanza di categorie incommensurabili, e che il loro uso portasse a un regresso infinito."
 - Il libro di Dennett affronta il problema della coscienza dal punto di vista materialistico; "*uno degli obiettivi di questo lavoro è spiegare la coscienza senza mai cedere al canto delle sirene del dualismo. Cosa c'è di tanto sbagliato nel dualismo? Perché è così osteggiato?... Un principio fondamentale della fisica è che ogni mutamento nella traiettoria di una particella è un'accelerazione che richiede una spesa di energia... tale principio di conservazione dell'energia... è*

apparentemente violato dal dualismo. Questo confronto tra la fisica standard e il dualismo è stato a lungo discusso dai tempi di Cartesio, ed è generalmente considerato come il difetto inevitabile del dualismo”

- Questa obiezione non si applica al modello H/J, il quale considera la coscienza causalmente effettiva, e completamente compatibile con tutte le leggi note della fisica, compresa la conservazione dell'energia.
- Secondo il materialismo Cartesiano, ci sarebbe *“un teatro centrale dove tutto accade all'unisono.. un luogo dove l'ordine di arrivo coincide con l'ordine di presentazione nell'esperienza, in quanto ciò che là accade è ciò di cui tu hai coscienza.”* Dennett confuta l'idea di “presentazione” come di un qualcosa di indipendente dallo spettatore. Egli si basa su un esperimento di **Kolers e Grunau**: due macchie di colore diverso vengono proiettate rispettivamente nei punti A e B, abbastanza vicini tra loro (p.es. 4°), per 150 ms ciascuna, con un intervallo di separazione di 40 ms; la prima macchia sembra che si sposti dal punto A al punto B cambiando di colore nel punto intermedio del segmento AB. Il modello Cartesiano, in cui c'è simultaneità tra evento oggettivo e evento fenomenico, non può spiegare un effetto simile. Dennett propone un modello alternativo, che rifiuta l'idea intuitiva di un flusso di coscienza unico.
- Secondo Stapp il problema è come si possano logicamente formare entità intrinsecamente unitarie all'interno di un sistema logico riduzionistico, dove ogni cosa è null'altro che l'aggregazione di parti più semplici. Ciò è di fatto impossibile, ecco perché James disperava di poter risolvere il problema della mente nell'ambito della fisica classica.
- occorre una cornice logica non strettamente riduzionistica che contempi una qualche operazione capace di formare degli interi. Un evento di Heisenberg risponde allo scopo, e il modello H/J è un controesempio esplicito alle convinzioni di Dennett. Questo modello non ha un teatro cartesiano, ma può conciliarsi con un flusso di coscienza del tipo descritto da James, e spiegare i risultati bizzarri di Kolers-Grunau, come quelli del *carousel* precognitivo di Grey Walter, o il coniglio cutaneo di Gerard e Sherrick, o il ritardo soggettivo di Libet.
- La fase di selezione fissa una struttura nel cervello che, nel caso di semplice attenzione, è una rappresentazione cerebrale del sé e dell'ambiente circostante, mentre nel caso di un intenzione ad agire, è quel tipo di rappresentazione che serve da schema al corso scelto dell'azione, e che la **teoria ideomotoria di James** identificava con la percezione della posizione e movimento corporei proiettata nel futuro.
- Questo modello dualistico
 - 1) È incompatibile con le leggi della fisica
 - 2) Considera la coscienza funzionalmente
 - 3) identifica ciascun evento cosciente, così come descritto da James, con l'immagine nel dominio psicologico di uno speciale tipo di evento di Heisenberg nel cervello.

- Come il famoso *gatto di Schrödinger* evolve in una sovrapposizione di sistemi macroscopicamente distinguibili (p.es. “gatto vivo” e “gatto morto”), così un uomo in una scatola nera, isolato dal mondo e da osservatori esterni, evolverà, sotto l’azione di leggi puramente meccaniche, in una sovrapposizione di sistemi, p.es. “uomo in piedi” e “uomo seduto”. Se però l’uomo osserva se stesso (non solo con la vista, ma in qualsiasi altro modo) e si verifica un *evento di detezione*, allora una delle due possibilità viene selezionata, mentre l’altra sparisce dall’universo delle possibilità. Il modello H/J postula che questo evento di detezione possa essere riportato indietro nel cervello al punto in cui la rappresentazione cerebrale di “uomo in piedi” (o di “uomo seduto”) era completamente formata. L’esperienza associata è un’immagine di questa rappresentazione, secondo un isomorfismo che è il nucleo del modello.
- Nell’esperimento immaginario del *gatto di Schrödinger*, un gatto è posto in una scatola con un meccanismo che rilascerà un gas velenoso se il decadimento di un nucleo radioattivo verrà rivelato da un dispositivo. In base all’equazione di Schrödinger, l’intero sistema, così come descritto dalla QT, evolverà in un sistema che è la sovrapposizione di due parti: in una parte il gatto è vivo, nell’altra è morto. Il fatto è che solo una di queste due possibilità apparirà all’esperienza di un osservatore.
- Nell’esperimento di *carousel precognitivo*, di *Grey Walter*, il soggetto vede una sequenza di *slides* proiettate su uno schermo, e sa che può, quando vuole, passare alla prossima slide premendo un pulsante. Ma quello che in realtà fa avanzare le slides è il segnale amplificato proveniente da un elettrodo impiantato nella corteccia motrice del paziente. I soggetti riferivano che appena stavano per premere il pulsante, ma prima che essi avessero deciso realmente di farlo, il proiettore avanzava da sé.
- Questo fenomeno, citato da Dennett, è spiegato da Stapp secondo il modello H/J: i movimenti volontari sono innescati dalla attualizzazione di una rappresentazione anticipata di quelle che ci si aspetta dovranno essere le rappresentazioni prossime future del corpo e del suo ambiente. Dato che il segnale nervoso della rappresentazione predittiva è intercettato dall’elettrodo, l’azione intenzionale viene come anticipata, e le rappresentazioni vissute di quello che in realtà avviene non collimeranno con questa aspettativa. Ciò verrà percepito come un’anomalia temporale nella rappresentazione che il soggetto ha di sé e del proprio ambiente. Il soggetto sarà sorpreso del movimento “premature” della slide, che è incapace di impedire.
- Il merito di una teoria scientifica non è tanto nella verificabilità delle sue singole proposizioni, quanto nella sua economia e consistenza interna, nel suo scopo, e nella coesività col resto della scienza. Il modello H/J, sebbene ai suoi primi esordi, risponde a tutti questi criteri.
- Il libro di *R. Penrose, The Emperor’s New Mind*, descrive la QT in dettaglio, soffermandosi sui due diversi modi attraverso cui può evolvere un sistema quantistico:

- a) il regolare sviluppo deterministico stabilito dalla legge quantistica fondamentale del moto, l'equazione di Schrodinger (il cosiddetto **processo unitario, U**);
- b) improvvisi e imprevedibili salti quantistici (**processo di collassamento o di riduzione, R**). Ciascun salto quantistico opera una "scelta" o "decisione" che attualizza una delle molteplici possibilità linearmente sovrapposte generate dal processo U .

Al termine del libro Penrose giunge alla conclusione principale: "*Io sto immaginando che l'azione del pensiero cosciente sia qualcosa di molto vicino a una scelta tra alternative in sovrapposizione lineare. Tutto questo ha a che vedere con la fisica sconosciuta che regola il confine tra U e R ...*".

- Stapp ritiene di portare avanti il lavoro di Penrose, facendo luce su aspetti precisi da lui non affrontati: possono i pensieri coscienti modificare il corso degli eventi cerebrali?
- Penrose cita molti esempi dalla matematica e dalla musica in cui un solo atto di pensiero sembra afferrare la totalità (p.es. la descrizione di Mozart sul suo modo di comporre, pag.534 *op.cit.*). Ai fisici che abbiano avuto a che fare con i problemi fondamentali della QT questa "globalità" dei pensieri coscienti richiama la globalità degli stati quantistici, evidente nel paradosso EPR. Penrose dice giustamente che questo paradosso non è come la correlazione tra eventi distanti comprensibili nel quadro delle idee classiche ordinarie sulla natura del mondo fisico. In quel mondo si può pensare a una situazione in cui due palle siano lanciate in direzioni opposte da un punto centrale, sapendo che una è bianca e l'altra nera. Se poi si trova che da un lato una delle due palle è bianca si può subito dedurre che l'altra è nera. Ma il paradosso EPR è tutt'altro! Qui il sistema composto di due particelle distanti agisce come una singola entità globale, nel senso che non è possibile imporre che "Quello che uno scienziato decide di fare da un lato di un sistema non può influenzare in alcun modo come il sistema risponderà nello stesso istante a una misura eseguita su di esso sul lato opposto". La profondità di questo collasso della concezione classica locale-riduzionistica dell'universo fisico non fu pienamente percepita da Einstein, Podolsky e Rosen: divenne chiara solo dopo che **John Bell** ebbe preparato la strada col suo famoso **Teorema di Bell**.
- Una buona parte del libro di Penrose cerca di sviluppare due argomenti che Stapp trova difettosi, o quantomeno non conclusivi:
 - a) Giustificare l'idea che **il processo R sia strettamente connessa alla gravità**. Penrose argomenta che se esistono buchi neri, e se lo spazio delle fasi scompare in essi, allora qualche altro processo, il processo R , deve creare una pari quantità di spazio delle fasi in regioni prive di buchi neri. Questa è un'assunzione goffa e artificiale. Anche se si prendesse per buona non spiega il processo R , dato che R *esclude* alcuni dei rami classici dello spazio delle fasi riempito dal flusso generato da U . Sicché R sembra semmai eliminare lo spazio delle fasi piuttosto che crearlo. Stapp crede che R non sia associato alla gravità

o alla *massa di Plank* (10^{-5} g), ma piuttosto che *i salti quantistici operino delle scelte tra possibilità distinguibili al livello macroscopico*. Certo, Stapp non risponde *perché* i salti si verifichino a questo livello, e *quale* tra le possibilità venga selezionata.

b) Supportare la tesi che *la coscienza umana debba avere accesso diretto al mondo platonico delle verità matematiche* astratte. Stapp classifica come mistica questa pretesa, e preferisce la nozione più “naturalistica” che vede le entità teoretiche scaturire dall’unione del mondo fisico delle particelle e dei campi, così come descritti dalla QT, col mondo dei pensieri e dei sentimenti, così come descritto dagli psicologi. L’intuito matematico, che trascende le possibilità di una TM in base al *Teorema di Gödel*, potrebbe essere semplicemente il risultato di una certa capacità di manipolazione visiva collaudata dall’esperienza.

- Nel libro di **Michael Lockwood**, *Mind, Brain, and the Quantum*, l’autore rileva tutti i pregiudizi e le nozioni datate che i filosofi della mente hanno sulla materia, che dopo l’avvento della QT rimane misteriosa quanto la mente. A differenza di Stapp e Penrose, Lockwood propone l’*interpretazione dei molti mondi* (*many-worlds interpretation*) o delle *molte menti* (*many-minds interpretation*) della QT. Secondo questa ipotesi il processo casuale **R** non sarebbe fisicamente reale, ma una specie di illusione, dovuta al fatto che la coscienza si separerebbe continuamente in ramificazioni non comunicanti: ciascun “Io” si divide in una molteplicità di altri “Io”, ciascuno con una diversa successione di esperienze.
- Anche **John Eccles** ha proposto un modello di interazione mente-cervello sulla base della QT, della quale sottolinea il non determinismo, senza però entrare veramente nel merito della teoria. Eccles ritiene necessario fare ricorso a un’“anima”, che conosce e controlla il cervello. I pensieri hanno struttura non fisica.

L’ONTOLOGIA DI HEISENBERG

- *Def.* Secondo la definizione ortodossa, data da Bohr e Heisenberg, la QT è un insieme di regole per fare previsioni sulle osservazioni ottenute in particolari condizioni sperimentali.
- Heisenberg non volle fermarsi a questa definizione epistemologica, e fornì una interpretazione ontologica del mondo fisico coerente con le implicazioni della QT. L’idea centrale è che gli atomi non sono cose “attuali”. Lo stato fisico di un atomo, o di un aggregato di atomi, è solo un insieme di tendenze oggettive al verificarsi di un certo peculiare tipo di “eventi attuali”. Inoltre l’evoluzione dinamica della natura non è un unico processo uniforme, come sostiene la CP, ma consiste di due processi. Uno è continuo, ordinato, deterministico, controllato da leggi matematiche che sono dirette generalizzazioni delle leggi della CP. Tuttavia

questo processo non controlla le cose attuali ma le propensioni, o tendenze oggettive, al loro verificarsi. L'altro processo fondamentale è una sequenza di salti quantici, che non sottostanno a nessuna legge nota della fisica, e sono descrivibili solo in senso statistico. Questi salti sono le cose "attuali" della natura, o gli eventi attuali di Heisenberg.

- Le nostre percezioni sensibili di un oggetto macroscopico composto da un gran numero di particelle possono essere rappresentate da un piccolo numero di variabili osservabili, quantitativamente tali che gli effetti quantistici sono irrilevanti. La nostra percezione diretta è limitata a queste variabili. Le proprietà di un sistema quantistico di dimensioni atomiche possono essere indagate solo con l'ausilio di dispositivi di misura. Il dispositivo deve essere in uno stato di equilibrio instabile, in modo tale che un piccolo segnale emesso dal sistema atomico possa innescare una serie di reazioni a catena, che portano a cambiamenti osservabili sullo strumento. L'attualizzazione di una delle configurazioni metastabili del sistema atomico sotto misura, con eliminazione di tutte le altre, è un evento attuale di Heisenberg.
- Quindi, le probabilità che figurano nelle leggi matematiche del moto non sono semplicemente le probabilità che la comunità degli osservatori umani osservi questo o quello stato metastabile (interpretazione ortodossa), ma le tendenze oggettive al loro verificarsi in natura. Questa distinzione è irrilevante ai fini di tutti gli esperimenti condotti finora.

IL MODELLO DI H.STAPP

- Stapp propone di considerare il cervello come un dispositivo che processa gli input sensoriali, costruisce una serie di possibili risposte corporee, ne seleziona una e controlla la sua esecuzione. Egli ipotizza l'esistenza nel cervello di uno ***schema di sé e del mondo*** (*self and world schema*) formato da:
 - a) uno ***schema corporeo*** (*body schema*) rappresentazione della disposizione attuale delle parti del corpo
 - b) uno ***schema del mondo esterno*** (*external world schema*) rappresentazione attuale dell'ambiente
 - c) uno ***schema delle credenze*** (*belief schema*).

Lo schema di sé e del mondo si aggiorna continuamente, passando nella parte "corrente" di uno ***schema storico*** (*historical schema*) quando è rimpiazzato da un nuovo schema, che si pone nella parte "proiettata" (*projected*) nel futuro.

Stapp osserva che:

- 1) Possiamo considerare il cervello un dispositivo di misura di Heisenberg, dove questi schemi sono entità osservabili, descrivibili e manipolabili classicamente.
- 2) I processi cerebrali sono chimici e quindi devono essere trattati dalla QT. Se una sinapsi scaricherà o meno dipende da eventi ionici. Esistono miliardi di sinapsi, quindi un numero enorme di pattern riverberanti metastabili di impulsi nei quali il cervello può evolvere, e che rappresentano gli schemi di cui sopra.

Molti di questi schemi metastabili verranno costruiti in sovrapposizione quantistica secondo l'equazione di Schrodinger, ma uno solo viene selezionato. Questo evento cerebrale è un evento cosciente e ogni evento cosciente è di questo tipo.

3) Esistono eventi di Heisenberg che non sono coscienti

- Il **tempo** è la dimensione della natura che permette l'evoluzione. Nella QT esistono due modi di evoluzione, e quindi due concetti di tempo:

1) il *tempo di processo*, o *tempo attuale*, è quello scandito dalla sequenza degli eventi attuali, cioè la successione discreta dei salti quantici nello *stato dell'universo* di Heisenberg

2) il *tempo di Einstein*, o *tempo virtuale*, è quello unito allo spazio nello spaziotempo. Esso è associato alle equazioni del moto di Heisenberg dell'evoluzione degli operatori che corrispondono alle quantità fisiche della CP. Questa evoluzione è virtuale nel senso che non riguarda le cose attuali stesse, ma solo le potenzialità e probabilità delle cose attuali.

-

Appunti integrati di

MECCANICA QUANTISTICA

- I **sistemi fisici** (S) possono essere classificati in base alle loro proprietà invarianti nel tempo, e lo **stato** del sistema a un dato istante è la specificazione completa delle sue proprietà che invece variano nel tempo.
- Nel linguaggio della meccanica quantistica (MQ) un **osservabile** (*observable*) è lo stesso che dire una **quantità fisica**.
- Lo **spazio di stato** (*state-space*) di un sistema S è l'insieme di tutti i suoi possibili stati, cioè tutti i modi fisicamente possibili di combinare i valori delle quantità che lo caratterizzano internamente. Lo spazio di stato di un sistema classico composto di n particelle si ottiene specificando i valori di $6n$ quantità a valori reali (tre componenti di posizione e tre momenti per ogni particella) e pertanto è uno spazio *6n-dimensionale*. Ciascuno stato possibile di S corrisponde a un punto nello spazio, e ciascun punto dello spazio corrisponde a uno stato possibile di S . La restrizione degli stati matematicamente descrivibili a quelli fisicamente possibili è

particolarmente opportuna nei sistemi quantistici dove non tutte le grandezze relative a S hanno valori definiti quando sono considerate simultaneamente (a causa p.es. del **principio di indeterminazione** di Heisenberg). Gli spazi di stato della MQ sono spazi vettoriali speciali, detti spazi di Hilbert, molto più strutturati internamente rispetto agli spazi di stato classici.

- Una **struttura** è un insieme di elementi su cui sono definite certe operazioni e relazioni. In una **struttura matematica** gli elementi sono oggetti matematici (numeri, vettori, insiemi) e le operazioni sono matematiche, appunto. Un **modello** è una struttura matematica usata per rappresentare una qualche struttura fisicamente significativa nel mondo. Dal punto di vista matematico ciò che principalmente distingue la MQ dalle teorie classiche che la precedono è che gli stati e le quantità hanno qui una più ricca struttura; essi formano famiglie con una più interessante rete di relazioni. Il cuore della MQ è contenuto negli **spazi di Hilbert**, strutture matematiche che rappresentano gli spazi di stato dei sistemi della MQ.
- Un vettore A , scritto $|A\rangle$, è un oggetto caratterizzato da una lunghezza $|A|$, e una direzione.
- Def. Uno **spazio vettore** V è un insieme di vettori chiuso rispetto all'addizione e alla moltiplicazione per una costante
- Def. $\langle A|B\rangle$ è un **prodotto interno** (*inner product*) su uno spazio vettore V sse
 - i) $\langle A|A\rangle = |A|^2$, e $\langle A|A\rangle = 0 \Leftrightarrow A = 0$
 - ii) $\langle B|A\rangle = \langle A|B\rangle^*$
 - iii) $\langle B|A+C\rangle = \langle B|A\rangle + \langle B|C\rangle$
- Def. Uno **spazio prodotto interno** è uno spazio vettore su cui è stata definita l'operazione di moltiplicazione tra vettori, e la **dimensione** di questo spazio è il numero massimo di vettori non nulli mutualmente ortogonali che esso può contenere.
- Una qualsiasi N -upla di vettori mutualmente ortogonali di lunghezza 1 in uno spazio vettore N -dimensionale costituisce una **base ortonormale** per quello spazio.
- Se $|A_1\rangle, \dots, |A_N\rangle$ è una base ortonormale, ogni vettore di V può essere espresso nella forma

$$|Q\rangle = q_1|A_1\rangle + q_2|A_2\rangle + \dots + q_N|A_N\rangle$$
 dove $q_i = \langle Q|A_i\rangle$ ($i = 1, \dots, N$) sono i coefficienti dell'espansione di Q nella base A .
- Def. Quando N è infinito, si usa denotare il vettore Q con la sua **funzione d'onda** $\psi(i)$, che esprime per ogni i , il coefficiente di espansione di Q relativo alla i -esima componente di base

$$\psi(i) = q_i = \langle Q|A_i\rangle \quad (i = 1, \dots)$$
- Def. Un **operatore** O è un'applicazione di uno spazio vettore V in se stesso:

$$O|Q\rangle = Q' \quad (Q, Q' \in V)$$
- Per un operatore lineare O si ha
 - i) $O(|A\rangle + |B\rangle) = O|A\rangle + O|B\rangle$

$$ii) \quad O(c|A\rangle) = c(O|A\rangle)$$

Proprio come ogni vettore in uno spazio N -dimensionale può essere rappresentato da una colonna di N numeri, relativamente a una base scelta per lo spazio, ogni operatore lineare sullo spazio può essere rappresentato con una matrice $N \times N$

$$O = \begin{vmatrix} O_{11} & O_{12} \\ O_{21} & O_{22} \end{vmatrix}$$

dove $O_{ij} = \langle A_i | O | A_j \rangle$, e $|A_n\rangle$ sono i vettori di base dello spazio.

- $|B\rangle$ è un **autovettore** di O con **autovalore** b sse $O|B\rangle = b|B\rangle$. Operatori diversi possono avere diversi autovettori, ma la relazione autovettore/operatore dipende solo dall'operatore e dai vettori in questione, non dalla particolare base in cui sono espressi; si tratta cioè di una relazione invariante rispetto al cambiamento di base.
 - *Def.* Un operatore lineare H è **Hermitiano** $\Leftrightarrow H$ ha solo autovalori reali.
 - *Def.* Uno **spazio di Hilbert** è uno spazio vettore su cui è definito un prodotto interno, e che è completo, cioè tale che ogni sequenza di Cauchy di vettori nello spazio converge a un vettore nello spazio. Tutti gli spazi prodotto interno a dimensione finita sono completi, quindi sono spazi di Hilbert.
 - Ogni sistema fisico è associato a uno spazio di Hilbert, ogni vettore unitario dello spazio corrisponde a un possibile stato puro del sistema, e ogni possibile stato puro del sistema corrisponde a qualche vettore dello spazio, rappresentato da una funzione d'onda ψ .
 - Gli operatori Hermitiani definiti sullo spazio di Hilbert associato a un sistema rappresentano quantità fisiche (*osservabili*), e i loro autovalori rappresentano i possibili risultati di una misura di queste quantità.
 - Quando due sistemi fisici S_1 e S_2 interagiscono, essi formano un sistema composto S_{12} e, tanto nella MC quanto nella MQ, esiste una regola per costruire, a partire dagli spazi di stato dei sistemi componenti (H_1 e H_2) lo spazio di stato del sistema composto (**prodotto tensore** $H_1 \otimes H_2$)
- Si dimostrano due fatti importanti:
- i) Dato che H_1 e H_2 sono spazi di Hilbert, anche $H_1 \otimes H_2$ è uno spazio di Hilbert
 - ii) Lo stato del sistema composto S_{12} non è univocamente definito dagli stati dei sistemi costituenti.
- La **dinamica** di un sistema quantistico può essere descritta in due tipi di contesto
 - 1) Noto lo stato del sistema all'istante t e le forze e i vincoli cui è soggetto, l'**equazione di Schrödinger** fornisce lo stato in qualsiasi istante successivo t' $U|Q_t\rangle \rightarrow |Q_{t'}\rangle$. Importanti proprietà di U sono il fatto che è un'equazione deterministica e lineare.
 - 2) **Contesto di misura**: l'atto di misurare l'osservabile H su un sistema in uno stato $|A\rangle$ determina il collassamento del sistema nell'autovettore di H che

corrisponde all'autovalore osservato (*postulato del collassamento*). Quale particolare autovettore di H venga interessato è questione di probabilità, e le probabilità sono date dalla *regola di Born*

$$Pr(b_i) = |\langle A | H = b_i \rangle|^2$$

- Nessuno ha ancora soddisfacentemente spiegato nei termini forniti dalla teoria, quali contesti siano contesti di misura e quali no. Inoltre, anche se esistesse questa distinzione, rimarrebbe il dubbio se esistano altre equazioni oltre quella di Schrödinger che governino la dinamica del sistema in un contesto non di misura.
- Le due più importanti proprietà degli spazi di Hilbert sono
 - P1) ogni vettore normalizzato dello spazio rappresenta uno stato possibile del sistema S , e nel caso esso sia esprimibile come somma di due autovettori di un osservabile H con autovalori distinti, allora esso non sarà un autovettore di H , ma sarà associato a un insieme di probabilità di avere un risultato o l'altro nella misura di H .
 - P2) Dato un operatore Hermitiano in uno spazio di Hilbert, ve ne sono altri, nel medesimo spazio, con i quali non ha autovettori in comune.
- Facciamo un'altra assunzione:
 - 4.1) Un sistema ha un valore b per un osservabile H sse il vettore che rappresenta il suo stato è un autovettore di H , con autovalore b .
- Da P2) segue che nessuno stato può essere un autostato per tutti gli osservabili, quindi nessun sistema può assegnare simultaneamente valori per tutte le quantità ad esso pertinenti.
- Dato un operatore Hermitiano H associato a un osservabile, la misura di questo osservabile sul sistema che si trova in uno stato $\langle Q |$, esprimibile come combinazione lineare di uno o più autovettori di H ,

$$|Q\rangle = q_1|A_1\rangle + q_2|A_2\rangle + \dots + q_N|A_N\rangle$$

darò un risultato a_i con probabilità $|q_i|^2$, essendo a_i autovalore di $|A_i\rangle$. Si dice che $|Q\rangle$ “collassa” in $|A_i\rangle$ se il risultato della misura è a_i .

- Esistono osservabili di S_{12} ($S_1 + S_2$) i cui valori non sono determinati dai valori degli osservabili di S_1 e S_2 .
-
- Resta un altro fatto sulla struttura matematica della teoria che è importante sapere. Non è una proprietà degli spazi di Hilbert, ma della dinamica, le regole che descrivono le traiettorie che il sistema segue nello spazio. Supponiamo di avere un sistema S e un dispositivo S^* che misura un osservabile A su S con valori $\{a_1, a_2, \dots\}$. Allora esiste un qualche stato di S^* (*ground state*), e un qualche osservabile B su S^* con valori $\{b_1, b_2, \dots\}$ (*pointer observable*) tali che, se S^* parte dal suo stato fondamentale (*ground state*) e interagisce in modo opportuno con S , e se il valore di A immediatamente precedente all'interazione è

a_1 , allora il valore di B immediatamente successivo è b_1 . Se invece se il valore di A immediatamente precedente all'interazione è a_2 , allora il valore di B immediatamente successivo è b_2 , ecc.

Ciò significa che S^* misura A.

Se rappresentiamo lo stato congiunto di $S+S^*$ mediante il vettore $|A=a_i\rangle_S$ $|B=b_i\rangle_{S^*}$, e se \rightarrow rappresenta la descrizione dinamica dell'interazione tra i due, allora dire che S^* è uno strumento di misura per A è come dire che le leggi dinamiche implicano che

$$|A=a_i\rangle_S |B=\text{ground state}\rangle_{S^*} \rightarrow |A=a_i\rangle_S |B=b_i\rangle_{S^*}$$

Intuitivamente, S^* è uno strumento di misura per A solo se esiste una qualche caratteristica osservabile di S^* che è correlata con i valori di A con i quali esso interagisce.

Ora, ci sono stati di S (troppi per contarli) che non sono autovettori di A, e se noi consideriamo cosa ci dice l'eq. di Schrödinger sull'evoluzione congiunta di S e S^* quando S evolve a partire da uno di questi, troviamo che lo stato congiunto dopo l'interazione è la sovrapposizione di autovettori di [A su S & B su S^*], l'osservabile i cui valori corrispondono alle assegnazioni congiunte di A e B. Ciò segue dalla linearità dell'equazione di Schrödinger, a prescindere da cosa B stia misurando su S.

Supponiamo per es. di far partire S^* dal suo *ground state*, e S dallo stato

$$1/\sqrt{2} |A=a_1\rangle_S + 1/\sqrt{2} |A=a_2\rangle_S$$

Segue dalle regole di composizione degli spazi di stato che lo stato combinato della coppia è

$$1/\sqrt{2} |A=a_1\rangle_S |B=\text{ground state}\rangle_{S^*} + 1/\sqrt{2} |A=a_2\rangle_S |B=\text{ground state}\rangle_{S^*}$$

Dal fatto che S^* è uno strumento di misura per A, e dalla linearità di U segue che il loro stato congiunto dopo l'interazione è

$$1/\sqrt{2} |A=a_1\rangle_S |B=b_1\rangle_{S^*} + 1/\sqrt{2} |A=a_2\rangle_S |B=b_2\rangle_{S^*}$$

Ciò tuttavia è inconsistente con la regola dinamica valida in un contesto tipo 2), la quale implica che lo stato congiunto dopo l'interazione dev'essere

$$|A=a_1\rangle_S |B=b_1\rangle_{S^*}$$

oppure

$$|A=a_2\rangle_s |B=b_2\rangle_{s^*}$$

cioè esisterebbe una probabilità $\frac{1}{2}$ che esso vada a coincidere con il primo, e una probabilità $\frac{1}{2}$ che esso vada a coincidere con il secondo.

Questo è il problema della misura (***Measurement Problem***) in MQ.

- Gli ***stati misti*** (*mixed states*) sono somme pesate di stati puri, e possono essere usati per rappresentare gli stati di insiemi i cui componenti sono in stati puri diversi, o stati di sistemi individuali di cui abbiamo solo una conoscenza parziale. Nel primo caso, il peso assegnato a un dato stato puro riflette la dimensione del componente dell'insieme che è in quello stato (e quindi la probabilità oggettiva che un arbitrario membro dell'insieme coincida con esso); nel secondo caso esso riflette la probabilità epistemica che il sistema in questione in cui lo stato è assegnato sia in quello stato.
- Se non vogliamo perdere la distinzione tra stati puri e misti, ci serve un modo per rappresentare la somma pesata di un insieme di stati puri (equivalentemente, le funzioni di probabilità associate), il che è diverso dalla somma pesata dei vettori che li rappresentano. Cioè abbiamo bisogno di un modo alternativo di rappresentare gli stati misti, oppure un modo uniforme di rappresentare sia gli stati puri che quelli misti pur preservando la distinzione tra di essi. C'è un tipo di operatore definito su spazi di Hilbert, detto ***operatore densità***, che soddisfa bene alla seconda soluzione. In questo caso, gli stati, puri o misti che siano, sono rappresentati non tanto da vettori, quanto da operatori densità. Sebbene gli stati misti possano essere usati per rappresentare la nostra ignoranza degli stati di sistemi che si trovano in realtà in uno stato puro o nell'altro, e sebbene ciò sia sembrato a molti un modo adeguato per interpretare le mescolanze in contesti classici, ci sono seri ostacoli nell'applicarli generalmente alle mescolanze in meccanica quantistica.

Tutto quello che è stato detto sugli osservabili, strettamente parlando, si applica solo al caso in cui i valori degli osservabili formi un insieme discreto; le finanze matematiche necessarie alla generalizzazione al caso di ***osservabili continui*** sono complicate.

- A dispetto degli straordinari successi predittivi, la MQ è stata penalizzata, sin dal suo esordio quasi 70 anni or sono, da alcune difficoltà concettuali. Detto in parole povere non è chiaro che cosa la MQ descriva in realtà.
- Dal fatto ampiamente accettato che lo stato di un qualsiasi sistema MQ sia completamente specificato dalla sua funzione d'onda, sembrerebbe che la MQ sia fondamentalmente il comportamento delle funzioni d'onda. Ma Schrödinger stesso, il padre della funzione d'onda, trovò infine impossibile crederlo.

- Tre punti di discontinuità tra la MQ e le teorie classiche
- 1) Due entità poste a distanza senza alcun mezzo per scambiarsi informazioni possono mostrare sorprendenti correlazioni nel loro comportamento
- 2) Un fotone può manifestarsi con proprietà tipiche di una particella o tipiche di un'onda, e rimanere in uno stato di ambiguità finché non venga fatta su di esso una misurazione. Il tipo di misura decide il modo in cui esso si manifesta
- 3) Se lo stato Q di un sistema S è una descrizione completa di S stesso, allora un osservabile che abbia un valore indefinito in Q è *oggettivamente* indefinito e non semplicemente *sconosciuto*
- 4) Il concetto di indeterminazione non è limitato al solo mondo atomico e subatomico. Si è scoperto che anche un sistema macroscopico può trovarsi in uno stato nel quale un osservabile ha un valore indeterminato.
- La **teoria delle variabili nascoste** ammette l'esistenza di proprietà dei sistemi non descritte dallo stato quantico. Pertanto non esiste un'indeterminazione oggettiva, né una probabilità oggettiva.

Nel 1964 John S. Bell del Laboratorio europeo per la fisica delle particelle del CERN ha dimostrato che esistono situazioni fisiche in cui nessun modello locale (in un senso ben preciso) a variabili nascoste può generare previsioni in accordo con quelle della MQ.

IL PROBLEMA DELLA MISURA IN MQ

- Ogni interazione tra sistema quantistico S e strumento di misura D non è più riducibile ad una registrazione da parte di D del valore della grandezza misurata su S , ma comporta una modifica del sistema S , non eliminabile né praticamente, né concettualmente.
- L'interazione può anche essere nulla, ma in tal caso non si compie alcuna osservazione e non si acquisisce alcuna conoscenza sulle proprietà del sistema. Se l'interazione è diversa da zero, essa non può essere resa arbitrariamente piccola, ma deve essere $\geq h/4\pi$ (**h costante di Planck**).
- Ogni interazione fisica tra D e S comporta sempre uno scambio di energia per un certo intervallo di tempo, oppure la cessione di una certa quantità di moto su una certa distanza spaziale. Le dimensioni fisiche di "energia per tempo" e di "quantità di moto per spazio" sono proprio quelle dell'**azione**, sono cioè quelle della costante di Planck. Il termine **interazione** usato per ogni processo misurazione quantistica è quindi vero letteralmente, dato che si ha proprio a che fare con lo scambio della grandezza fisica *azione* fra D e S .
- Date le dimensioni atomiche di S , la misura comporta una inevitabile perturbazione di S stesso. Questo carattere perturbativo della misurazione in

microfisica comporta la necessità dell'abbandono dello schema classico della doppia registrazione da parte dell'apparato prima e dell'osservatore poi, a favore del nuovo schema :

$$S \leftrightarrow D \rightarrow O$$

che indica esplicitamente la natura simmetrica e non più unidirezionale del processo di interazione tra apparato misuratore D e sistema atomico misurato S . L'interazione tra apparato D e osservatore O continua ad essere indicata come una registrazione da parte di O del risultato da lui osservato su D . Si tratta di uno schema che non è comunque universalmente accettato e sono state avanzate, come vedremo, proposte più radicali, che sostengono la natura simmetrica anche dell'interazione tra O e D , e la conseguente possibilità di una modifica, da parte della mente o della coscienza di O , dello stato fisico di D e, attraverso questo, dello stesso S .

- In MQ, la descrizione matematica del processo di interazione tra D e S comporta tuttavia una serie di gravi difficoltà (***problema della misurazione***). Le radici logiche del problema vanno ricercate nella coesistenza, all'interno del formalismo, di due differenti (e tra loro incompatibili) modalità di evoluzione della funzione d'onda associata allo stato di S . Da un lato abbiamo infatti l'***equazione di Schrödinger*** che regola in modo deterministico l'evoluzione temporale dei sistemi fisici inosservati, dall'altro la ***descrizione probabilistica*** delle osservazioni o misurazioni effettuate per determinati osservabili dei sistemi fisici stessi, senza che venga tuttavia formulata una specificazione non ambigua di condizioni precise e reciprocamente esclusive per questi due differenti tipi di evoluzione.
- Infatti se si applicasse l'equazione di Schrödinger alla descrizione del processo di misura interpretandolo come un qualsiasi altro processo di interazione tra due oggetti fisici, corrispondenti al sistema misurato S e all'apparato misuratore D , ogni volta che S è in uno stato di sovrapposizione di due o più stati microscopici, lo stato finale del sistema complessivo $S+D$ risulterebbe, a causa della linearità dell'equazione di Schrödinger, ancora uno stato di sovrapposizione, in questo caso di differenti stati macroscopici. Ciò significa che O noterebbe D registrare simultaneamente tutti i differenti risultati previsti dal formalismo quantistico!
- Per evitare questa conclusione contraddittoria, nella QT è stato introdotto un assioma, che stabilisce che l'equazione di Schrödinger non si applica alla descrizione dei risultati delle nostre misurazioni o delle nostre osservazioni. Si tratta del cosiddetto ***postulato di riduzione della funzione d'onda***, secondo il quale il sistema misurato S è sempre in uno stato ben definito al termine del processo di misura.
- Distinguiamo quattro gruppi di teorie della misurazione, a seconda del livello a cui dovrebbe avere luogo il processo di riduzione d'onda, il passaggio cioè da una sovrapposizione di diversi stati fisici a un preciso stato finale.

La **PRIMA RISPOSTA**, contenuta nella teoria della misurazione di Bohr (scuola di Copenhagen), afferma che la riduzione della funzione d'onda avviene a livello

dello strumento di misura D . Questa ipotesi riconosce, quindi, allo strumento, uno status particolare rispetto a quello degli ordinari sistemi macroscopici, le cui interazioni con i microoggetti continuano ad essere descritte dall'equazione di Schrödinger. Nel caso di Bohr lo schema del processo di misura è del tipo

$$S \leftrightarrow D \rightarrow O$$

e il dominio di validità del formalismo quantistico si estende ai sistemi microscopici (S_A) e macroscopici (S_M) escludendo gli strumenti di misura.

La **SECONDA RISPOSTA**, contenuta nella teoria della misurazione di **J.von Neumann** (*"The Mathematical Foundations of Quantum Mechanics"*, Princeton University Press, New Jersey, 1955), assume che il processo di riduzione della funzione d'onda si verifica a livello della coscienza dell'osservatore. Il dominio di validità si estende quindi a tutti gli oggetti fisici, compresi gli strumenti di misura, escludendo la coscienza dell'osservatore che, concepita come agente esterno al mondo fisico, sarebbe la responsabile della riduzione d'onda. Il processo è rigorosamente simmetrico,

$$S \leftrightarrow D \leftrightarrow O$$

Ciò che dovrebbe avvenire in realtà è un trasferimento a catena di sovrapposizione di stati. La particella osservata "riversa" la propria sovrapposizione sull'apparato di misura, l'apparato di misura sugli organi di senso e questi sul cervello dell'osservatore. La catena di "condizionamenti reciproci" non avrebbe mai fine se non fosse per l'entrata in scena della coscienza (o "facoltà di introspezione") dell'osservatore la quale, per le sue caratteristiche di "irriducibilità", permette a questo di conoscere il proprio stato, troncando la catena delle sovrapposizioni e consentendo la "misura" del microoggetto.

La teoria di von Neumann, ripresa in tempi recenti da **E.P. Wigner** (*Remarks on the Mind-Body Question, Symmetries and Reflection*, Bloomington, Indiana University Press, 1967), è oggi conosciuta in tre diverse versioni :

- a. **Interazionista**. L'atto di misura modifica, attraverso il cambiamento del valore della funzione d'onda, il mondo fisico;
- b. **spiritualista** (o mentalista). Non esiste una realtà fisica oggettiva, ma soltanto le idee o gli stati mentali del soggetto che osserva (teoria delle osservazioni di Wigner);
- c. **solipsista**. Soltanto il mio io, e non semplicemente la coscienza di un qualsiasi altro osservatore è in grado di produrre la riduzione della funzione d'onda (paradosso dell'amico di Wigner).

Nessuna di queste tre varianti (ad eccezione, forse, della "a") sembra comunque fornire una soluzione soddisfacente; riducendo il processo di misura a una sorta di evento psicofisico.

La **TERZA RISPOSTA**, contenuta nella teoria della misurazione di *Hugh Everett* (B.De Witt e N.Graham, *The Many-Worlds Interpretation of Quantum Mechanics*, Princeton University Press, Princeton, New Jersey, 1973), afferma che la riduzione della funzione d'onda non ha in realtà mai luogo e che tutte le possibilità previste dal formalismo si realizzano simultaneamente in differenti ramificazioni dell'universo. Nonostante questo suo carattere paradossale la teoria di Everett ha una sua consistenza interna e l'importante vantaggio di estendere dominio di applicazione della QT agli oggetti macroscopici S_M , agli oggetti microscopici S_A , agli strumenti di misura D e agli stessi osservatori O , tutti concepiti come elementi della realtà fisica, mentre lo schema del processo di misura, che tiene conto del processo di ramificazione dell'universo, può essere così esemplificato

$$\begin{array}{l} S_{A1} \leftrightarrow D_1 \rightarrow O_1 \\ S_{A2} \leftrightarrow D_2 \rightarrow O_2 \\ \dots \\ S_{An} \leftrightarrow D_n \rightarrow O_n \end{array}$$

La più seria difficoltà della teoria di Everett è legata al fatto che essa non è in grado, in determinate situazioni fisiche, di spiegare il processo di ramificazione senza fare ricorso ad un agente esterno al mondo fisico, ricadendo quindi in una prospettiva simile a quella della teoria "dell'osservatore cosciente" di von Neumann.

La teoria della misurazione di Hugh Everett successivamente ripresa e ampliata da Bryce De Witt è anche nota col nome di **teoria degli Universi Paralleli** o **dei Molti Mondi**. In essa tutte le possibilità quantistiche alternative assumono realtà oggettiva e coesistono in parallelo l'una accanto all'altra. Ogni volta che si compie la misurazione di un microoggetto l'universo si scinde in due universi fisicamente reali e "identici" (la sola differenza sta nel diverso "stato" della particella osservata).

Può essere che il buonsenso si ribelli all'idea di un cosmo che si divide in due ogni volta che si esegue una misurazione, ma la teoria, da questo punto di vista, non è meno accettabile della stessa interpretazione di Copenhagen.

Un'alternativa alla suddivisione fisica degli universi è stata proposta da David Albert e Barry Loewer della Columbia University. Nella loro **teoria delle Molte Menti** i due studiosi americani ipotizzano che a ciascun osservatore o sistema fisico "senziente" sia associato un insieme infinito di stati mentali che esperiscono i diversi esiti possibili delle misurazioni quantistiche. La famiglia delle scelte latenti nell'equazione di Schrödinger corrisponde così alla miriade di esperienze subite da queste menti e non a un'infinità di universi paralleli materiali.

La **QUARTA RISPOSTA** è che la riduzione si verifica nel passaggio dal livello microscopico a quello macroscopico. Diversi tentativi sono stati elaborati in questa direzione. Ci limitiamo a citarne tre:

- a. le **Regole di Superselezione di Landau**, le quali vietano la sovrapposizione di stati macroscopici;
- b. le **teorie della misurazione di Loinger-Prosperi** e collaboratori (G.M.Prosperi, A.Loinger, A.Daneri, *Nuclear Physics* 33, 1962; G.M.Prosperi, *Models of the measuring process and of macrotheories*, Marburg, 1973), di Ludwig (G.Ludwig, *Die grundlagen der Quantenmechanik*, Berlino, 1954), di Prigogine e collaboratori (I.Prigogine, C.George, L.Rosenfeld, *The macroscopic level of quantum mechanics*, Det. Klg. Danske Vidensabernes, Meddelsen, 1972; I.Prigogine, *From Being to Becoming*, Freeman, S.Francisco, 1980). Secondo questi studiosi lo schema del processo di misura è simile a quello di Bohr, ma il dominio di applicazione del formalismo quantistico viene ristretto ai soli sistemi atomici (S_A), assumendo che la descrizione dei fenomeni macroscopici, e tra essi anche quella della interazione, richieda l'elaborazione di una nuova macrodinamica quantistica, in mancanza della quale è opportuno fare ricorso (per la descrizione del processo di misura) a teorie classiche o semi-classiche come la meccanica stocastica;
- c. la **teoria di Ghirardi-Rimini-Weber (GRW)** (C.Ghirardi, A.Rimini, T.Weber, *Physics Review D* 34, 1986) la quale afferma che la riduzione della sovrapposizione degli stati avviene spontaneamente nella transizione dalle dimensioni microscopiche alle dimensioni macroscopiche di D . Nell'ambito di questa teoria, al passar del tempo, la funzione d'onda di una particella si "espande". Durante l'espansione vi è la probabilità che l'onda urti "qualcosa sullo sfondo" e d'un tratto diventi localizzata. Per le particelle singole la probabilità di un tale urto è molto bassa, circa 1 ogni 100 milioni di anni, ma alle dimensioni degli strumenti di misura le probabilità del crollo della funzione d'onda salgono a 1 ogni 100 picosecondi.

- Per quanto le teorie sopra esposte sembrano realizzare un sostanziale progresso nei confronti dell'interpretazione data da Bohr, sia sul piano formale (per la rigorosa trattazione da esse offerta del processo di interazione tra S_A e D), sia sul piano fisico ed epistemologico, dato che la distinzione tra microscopico e macroscopico appare assai più fondata di quella tra sistema fisico e strumento di misura, esse vanno incontro ad alcune gravissime difficoltà, che impediscono (almeno allo stato attuale) di considerarle una reale soluzione al problema della riduzione della funzione d'onda nel processo di misura.

EQUAZIONE DI SCHRODINGER

- Nel 1926 il matematico austriaco Erwin Schrödinger scoprì un'equazione che permise e permette tutt'oggi di descrivere compiutamente le proprietà ondulatorie della materia. Questa equazione rappresenta lo stato di un atomo o di una particella isolata mediante una funzione d'onda, raffigurabile come un'onda

progressiva per una particella in moto nello spazio e come un'onda stazionaria per uno stato quantico di un atomo.

- Schrödinger chiamò tale funzione $\psi(\mathbf{r}, t)$.

L'equazione di Schrödinger nella sua forma classica:

$$\partial^2 \psi / \partial x^2 + \partial^2 \psi / \partial y^2 + \partial^2 \psi / \partial z^2 = (1/v^2) \partial^2 \psi / \partial t^2$$

dove v è la *velocità di propagazione*, t il tempo.

Il successo dell'equazione di Schrödinger nello spiegare la struttura atomica, e nel prevedere con estrema precisione un'incredibile serie di dati osservati, fu straordinario e con pochi precedenti nella storia della scienza.

Vediamo ora come questa equazione trova applicazione nell'ambito della QT. L'estensione della ψ al dominio della meccanica quantistica è dovuta essenzialmente a Max Born.

Secondo Born ad ogni particella deve essere associata un'onda descritta dalla funzione ψ . Tale funzione non deve però avere un significato fisico diretto, ma il quadrato del suo modulo deve rappresentare la densità di probabilità di presenza della particella in un determinato spazio.

L'interpretazione probabilistica dell'equazione di Schrödinger costituisce un aspetto peculiare della QT, poiché con essa le leggi rigidamente deterministiche della MC sono sostituite da leggi probabilistiche.

All'interno della MQ l'interpretazione probabilistica della ψ fa emergere un'ulteriore aspetto : la "sovrapposizione degli stati".

Ogni sistema quantistico che può essere contemporaneamente in uno stato C o stato B sarà quindi descritto da una funzione d'onda totale:

$$\psi(P) = P_C \psi(C) + P_B \psi(B)$$

e la probabilità di trovarlo nello stato B sarà $|P_B|^2$. Il fatto che la probabilità di un sotto-stato sia il quadrato del coefficiente lo si deve al formalismo, il quale richiede che la probabilità di presenza sia il quadrato di ψ . Perché la particella così descritta sia osservata occorre infine che la probabilità totale sia uguale all'unità; pertanto:

$$|P_B|^2 + |P_C|^2 = 1$$

La MQ offre esempi molto significativi di sovrapposizione di stati; uno di questi è la durata della vita del mesone K^0 . La durata della vita di K^0 può essere 10^{-7} oppure 10^{-10} sec, la qual cosa sembra violare il principio che prevede che una particella debba sempre presentare la medesima probabilità di disintegrazione. Applicando la teoria della sovrapposizione degli stati possiamo dire che la durata della vita di K^0 è la sovrapposizione di due stati K_1^0 (10^{-7}) e K_2^0 (10^{-10}). Il mesone può così

manifestarsi sotto una forma o l'altra a seconda del momento scelto per la sua osservazione. La vita di K^0 in definitiva si consuma per metà della sua durata in uno dei due stati. Questa situazione trova esemplificazione nella formula:

$$|K^0|^2 = \frac{1}{2} |K^0_1|^2 + \frac{1}{2} |K^0_2|^2$$

Ricapitoliamo ora quanto sino a questo punto esposto.

L'ampiezza dell'onda di un particella deve esser sempre rappresentata dalla sovrapposizione di due ampiezze ψ_1 e ψ_2 , lo stato di un sistema a un istante qualsiasi risulterà dato da $\psi = \psi_1 + \psi_2$ e poiché tale stato è proporzionale al quadrato dell'ampiezza dell'onda, scriveremo:

$$|\psi|^2 = |\psi_1 + \psi_2|^2 = |\psi_1|^2 + |\psi_2|^2 + |2\psi_1\psi_2|$$

Occorre precisare che il termine incrociato è nullo se entrano in gioco stati (irriducibili o puri) come quelli di particella-antiparticella o mesone $K^0_1 - K^0_2$, ovvero stati che non possono (assolutamente) esistere insieme. In questo caso l'ampiezza della funzione d'onda risulterà:

IL PARADOSSO EPR

- Il paradosso EPR è uno dei piu' dibattuti *Gedankenexperiment* nel campo della MQ. In particolare ci troviamo nel campo della MQ, così come e' regolata nell'interpretazione ortodossa (o di Copenhagen).
- L'**interpretazione di Copenhagen** assume i seguenti postulati:
 - 1) Lo stato di ogni sistema fisico S e' rappresentato da un vettore di stato v che appartiene ad uno spazio di Hilbert X
 - 2) Ogni misurazione e' formalizzata attraverso un operatore Hermitiano A
 - 3) Gli scalari che si ottengono dall'applicazione di un certo operatore A ad un insieme di elementi v_n dello spazio di Hilbert X sono i corrispondenti autovalori a_n (ovvero i risultati della misurazione)
 - 4) L'evoluzione di un vettore di stato v nel tempo e' regolata dalla relazione: $v(t) = T(t) * v(0)$ essendo $v(t)$ lo stato del sistema al tempo t generico e $v(0)$ lo stato iniziale. $T(t)$ è l'operatore di evoluzione, anch'esso dipendente da t .
 - 5) Una serie di misurazioni A sul medesimo insieme di sistemi uguali S fornisce un risultato medio m (postulato che costituisce il nocciolo dell'interpretazione frequentista)
 - 6) La misurazione A di un certo sistema fisico definito dal vettore di stato v fornisce un certo risultato r che e' correlato all'insieme di autovalori corrispondenti alla coppia (A, v) . All'atto stesso della misura v diviene r (il che si esprime dicendo che la funzione d'onda associata al sistema e' collassata ad uno scalare definito r).

Chiaramente dei precedenti punti l'ultimo e' quello più difficile da afferrare. In effetti il postulato (6) afferma che dato un certo esperimento esiste tutta una gamma (continua) di risultati possibili. Ad ogni risultato e' associata una certa probabilità di essere effettivamente rilevato. All'atto della misurazione questa incertezza probabilistica si risolve nella certezza di un numero ben determinato. L'osservazione del fenomeno diviene quindi parte fondamentale della medesima realtà che si vuol misurare. Inoltre la probabilità non nulla associata a certi esotici stati del sistema ci pone di fronte a situazioni concettualmente difficilmente interpretabili. Il tipo di dilemmi legati all'interpretazione probabilistica del mondo atomico e subatomico sembra svanire a scale macroscopiche. In effetti su "grandi" scale il comportamento dei sistemi fisici ricalca la descrizione deterministica classica. In particolare esiste un risultato teorico (Teorema di Ehenfest) che afferma che il "centro della distribuzione di probabilità " legata ad un certa grandezza osservabile propria di una particella si "muove" come un punto materiale classico nelle stesse condizioni. Il precedente risultato rende conto del Principio di Corrispondenza che possiamo vedere come una delle linee guida fondamentali nello sviluppo della MQ. In particolare questo principio afferma che deve esistere una corrispondenza formale delle relazioni che sovrintendono ai sistemi fisici macroscopici e quelle proprie della MQ. Seguendo questa linea situazioni particolari come l' Effetto Tunnel possono interpretarsi classicamente alla luce della teoria delle onde. Tuttavia e' necessario ricordare come la MQ si sia sviluppata sull'onda dei successi interpretativi della fisica dei quanti. Il mondo delle particelle e' quindi eternamente in bilico tra i due schemi interpretativi: corpuscolare ed ondulatorio. L'apparente problema sulla natura della materia e' ,a questo stadio, risolto dal Principio di Complementarità . Detto principio fonde due aspetti a prima vista incompatibili.

Una particella e' (per definizione) un oggetto localizzato nello spazio caratterizzato da grandezza quali impulso ed energia. Contrariamente un'onda e' nello spazio diffusa e le sue caratteristiche sono del tipo frequenza ed ampiezza.

Bohr propose di risolvere il problema con un radicale cambio paradigmatico sintetizzato dalle parole :

"...un fenomeno fisico non e' tale finche' non viene osservato"

"E' sbagliato pensare che il compito della fisica sia dire come la Natura e'.

La fisica verte su cio' che della natura possiamo dire."

Ecco quindi che , stante le sostanziali differenze di preparazione degli esperimenti, la decisione tra un modello interpretativo e l'altro e' compiuta all'atto stesso dell'osservazione. E' quindi la misurazione un nuovo ente che viene a far parte imprescindibilmente dello stesso fenomeno fisico sul quale si compie. Il particolare tipo di esperienza selezione tra onda e corpuscolo.

Cio' d'altronde riflette il carattere matematico di anticommutativita' che allo stesso Heisenberg (padre del principio di indeterminazione) sembrava assurdo pur scaturendo in maniera esplicita e rigorosa dalla sua stessa descrizione matematica. In ogni caso anche senza ricorrere all'algebra delle matrici di Heisenberg e semplicemente attraverso un esperimento di diffrazione e' possibile ricavare le relazioni di indeterminazione tempo-energia quantità di moto-posizione e con esse l'anticommutativita' dell'entrata in gioco delle grandezze osservate.

2° L'obiezione einsteiniana

Classicamente la posizione di Einstein in merito alla MQ (ed in particolare all'interpretazione che delle sue conseguenze dava Bohr) e' sintetizzata dalle sue stesse parole :

"Dio non gioca a dadi"

L'interpretazione probabilistica del mondo microscopico costituiva per Einstein la dimostrazione dell'ignoranza umana nella comprensione della Natura. Solamente dalla limitatezza del pensiero umano nasceva la necessita' di dare una descrizione che prevedesse il caso. In realtà Einstein pensava che tutto il sensibile fosse soggetto al determinismo. Ogni aspetto naturale doveva poter essere descritto in maniera Newtoniana. Sostanzialmente l'universo doveva intendersi come un enorme meccanismo nel quale ogni ingranaggio era perfettamente controllabile, misurabile e prevedibile nei suoi futuri movimenti. Per dimostrare l'intrinseca incompletezza della MQ (ed in particolare nei suoi fondamenti legati al principio di indeterminazione) Einstein scelse la via a lui consueta degli esperimenti concettuali. Uno di questi fu sviluppato insieme a Podol'skij e Rosen e presentato in Physical Review (47) del 1935. Oggi e' universalmente conosciuto come paradosso EPR.

Esso si basa essenzialmente sul Principio di Non Località . La posizione che Einstein aveva in merito al significato di una teoria fisica e' sintetizzata dalla sua convinzione di una realtà indipendente dall'osservatore. Chiaramente Bohr ed Einstein occupano due posizioni diametralmente opposte. Nell'articolo di cui ci occupiamo E.P.R. premettono i seguenti punti :

1. Una teoria fisica e' completa qualora (definizione) ogni elemento della realtà descritta abbia corrispondenza con un elemento teorico
2. Un elemento costitutivo della realtà fisica e' tale solamente nel caso in cui la teoria ne possa prevedere con certezza il valore senza che l'osservazione possa influenzarlo.

In particolare il gruppo EPR si sofferma sul concetto di Separabilita' che (nell'accezione einsteiniana) significa garantire l'impossibilita' di influenza istantanea tra due sistemi fisici distanti conformemente a quanto enunciato nei principi della Teoria della Relatività Ristretta (TRR) in merito al Limite di Velocità .

Consideriamo due sistemi fisici S1 ed S2 che interagiscano tra di loro. Lo stato dei sistemi complessivo S12 sara' descritto da un certo vettore di stato V_{12} la cui evoluzione temporale sara' regolata da un certo operatore di evoluzione $T=T(t)$. Ora separiamo i due sistemi (spazialmente). Per Heisenberg sappiamo che non tutte le grandezze caratterizzanti i sistemi S_i ($i \in \{1,2\}$) possono essere

contemporaneamente specificate. Scegliamo quindi per ognuna delle coppie di variabili coniugate dal principio di indeterminazione (nelle sue varie forme) un rappresentante. Infine avremo una certa quantità di grandezze misurate ammettiamo per S1. Da queste informazioni e da, analoghe, osservazioni sul sistema S2 possiamo desumere l'intera funzione d'onda Ψ di S2 dando un valore alle grandezze che venivano escluse dalle misurazioni fatte per S1. Tuttavia se avessimo deciso di specificare variabili diverse per S1 avremmo ricavato un diverso vettore di stato per S2. Poiché non possiamo accettare (ricordiamo le premesse di EPR) la possibilità di influenze di S1 su S2 (a distanza) dobbiamo desumere da questa esperienza l'incompletezza (einsteiniana) della MQ.

In maniera più spicciola il paradosso EPR può anche essere riassunto come segue. Consideriamo un sistema a spin nullo e supponiamo che esso emetta due fotoni. Per quanto prevede la MQ lo spin dei fotoni può essere solamente (spin quantizzato) del tipo Spin su oppure Spin giù. Fin quando lo spin del fotone (come di ogni altra particella e qualsiasi sua grandezza caratteristica) non viene osservato direttamente il fotone esiste in una sovrapposizione di spin. Per ognuna delle ipotesi possibili è assegnata una probabilità di osservazione. Se trascorso un tempo T dall'istante di emissione dei fotoni si osserva lo spin di uno dei due (diciamo spin su) poiché il momento di spin totale è una grandezza conservata si è contemporaneamente determinato (spin giù) lo spin del secondo fotone.

La risposta di Bohr fu coerente con le sue posizioni precedenti per quanto, tradizionalmente, non chiarisse definitivamente il problema. Egli fece notare come nell'esperimento EPR mancasse un'osservazione diretta dello stato di S2. In particolare il vettore di stato del secondo sistema veniva specificato per via teorica da EPR e non certo misurandone le grandezze caratteristiche. Mancava quindi la fondamentale misurazione diretta senza la quale, per Bohr, il fenomeno fisico non poteva essere considerato tale. In maniera più precisa e bene sottolineare che quando entrano in gioco grandezze dell'ordine di h (costante di Planck) allora è necessario fare i conti con i disturbi implicati dall'osservazione diretta. In questo senso realtà ed osservatore non possono considerarsi completamente separati.

Per il fenomeno EPR sembrano possibili solamente le seguenti due ipotesi:

1. Il vettore di stato del secondo sistema subisce un collasso istantaneo non locale per mezzo dell'osservazione dello stato del primo sistema. In questa ipotesi non viene contraddetto il principio di indeterminazione ma si presume un'interazione a distanza "tramite fantasmi" (come Einstein stesso disse) tra le particelle
2. La MQ è incompleta, fornendo solamente una descrizione probabilistica della realtà essa sfugge alla spiegazione di certi fenomeni (come l'EPR appunto). In effetti la MQ non riesce a tener conto degli aspetti che (come nel caso degli spin fotonici) sono legati una volta per tutte al momento stesso dell'inizio dell'esperimento.

La seconda ipotesi diede il via a quelle che classicamente sono dette **Teorie a Variabili Nascoste** (TVN) (tra di esse e' sicuramente da ricordare quella elaborata da Bohm). In queste teorie la punta probabilistica dell'iceberg interpretativo quantistico emerge come ignoranza umana nella capacita' di descrizione della parte soggiacente dell'iceberg stesso e che riguarda proprio certe variabili (VN) che a noi sfuggono. Principalmente le formulazioni TVN precisano l'esistenza di VN Interne ed Esterne. Le prime sono legate alla struttura delle particelle. Le seconde rendono conto delle perturbazioni presenti nelle regioni abitate dalle particelle. Nelle TVN si fa generalmente l'ipotesi di localita' escludendo quindi fenomeni EPR a priori. Si deve a Bell l'importante **teorema di selezione tra MQ e TVN**. In particolare esso permette di discriminare tra esperimenti interpretabili tramite TVN e MQ. Il fatto notevole e' che finora (se si eccettuano i casi in cui l'apparecchiatura di laboratorio favoriva un determinato corso delle cose) le previsioni fatte dalla MQ si sono rivelate sempre in accordo con i risultati sperimentali. Ai risultati di Alain Aspect (Università di Parigi) si sono recentemente aggiunte ulteriori riprove dell'esattezza delle previsioni della MQ. A questo proposito illustriamo un esempio sperimentale tipico (dovuto a R.Y.Chiao , P.G. Kwiat , A.M. Steinberg). In questa esperienza si analizzano i percorsi intrapresi da due fotoni emessi per suddivisione da parte di un certo cristallo. Essi hanno la possibilità di scegliere una deviazione (interferometro) lungo un percorso imposto che li porta dalla sorgente alla destinazione lungo due strade separate. La scelta effettuata discrimina tra una strada più breve ed un'altra meno diretta. I fotoni possono quindi uscire dai 2 interferometri attraverso 2 uscite distinte, che chiameremo destra e sinistra. Preliminarmente si e' osservato che la probabilità di uscire a destra oppure a sinistra sia la medesima per le due particelle. Intuitivamente si sarebbe portati a pensare che i risultati dell'osservazione siano del tutto casuali e che ,in particolare, non esistano relazioni tra l'uscita scelta dal fotone 1 e quella scelta dal fotone 2. Banalmente non e' così'. In effetti fissate determinate lunghezze della deviazione i 2 fotoni scelgono la medesima uscita. Oppure ,scegliendo altre lunghezze, le uscite sono sempre opposte. Il fatto che si possa passare (senza continuita') dal primo al secondo caso pone forti dubbi sull'idea che i due fotoni siano previamente correlati nella scelta. Infatti cio' equivarrebbe a dire che le due particelle sanno in anticipo il percorso che le aspetta. Dobbiamo concludere che , malgrado nessuno dei 2 fotoni sappia quale strada intraprenderà , una volta intrapresa esso conosce (non localita') istantaneamente la scelta del compagno. Per studiare le correlazioni suindicate si e' pensato di procedere al conteggio dei fotoni emergenti contemporaneamente dalla medesima uscita ed in coincidenza. Il conteggio in coincidenza equivale a richiedere che le particelle abbiano intrapreso il medesimo percorso , scegliendo entrambe la strada lunga oppure quella corta. Il caso di coincidenza e' interessante perche' da luogo ad interferenza , cosa che non accade se i percorsi fatti sono diversi essendo consequenzialmente differenti i tempi di uscita. Il fatto che queste due diverse "storie" possano coesistere implica per i due fotoni di aver percorso entrambe le vie possibili. Cio' evidentemente e' possibile solamente se i due fotoni sono stati emessi due volte , una per ogni strada possibile , e chiaramente

in tempi diversi visto il differente percorso da effettuare. Dall'analisi delle frange di interferenza che si osservano alla destinazione si rileva un "contrasto" superiore al 71% (in effetti vicino al 90%). Questo dato può essere utilizzato dal Teorema di Bell per dimostrare che un tale effetto non può essere descritto da una qualsivoglia teoria del tipo TVN. Malgrado cio' questo esperimento (insieme agli altri condotti con metodologie diverse ed a sostegno dell'interpretazione della MQ) non contraddice la TRG. La contraddizione esisterebbe solamente nel caso in cui si potesse decidere a priori la scelta (destra , sinistra) effettuata dall'un fotone o dell'altro. Nell'esperimento semplificato dello spin ciò significherebbe determinare prima dell'emissione fotonica lo spin di un fotone in modo tale che si abbia la determinazione dello spin per la seconda particella a distanza con un trasporto d'informazione superluminale. Ma (purtroppo , per fortuna ?) così non e'. La scelta che il fotone 1 effettua e' del tutto casuale (nel senso che non e' prevedibile stante le conoscenze che abbiamo) e conseguentemente non vi e' alcun trasporto di informazione possibile.

3° Conclusioni

Il paradosso EPR (che come si e' visto nasce essenzialmente da una scorretta interpretazione della MQ) mette in luce un aspetto importante. Come coniugare la MQ con una visione maggiormente vicina al senso comune della fisica ? Dobbiamo definitivamente rinunciare ad un simile progetto oppure e' ancora il caso di lavorare per far si che la nostra visione del mondo non sia solamente algoritmica ?. E' evidente allo stato attuale che la visione fornita dalla MQ e della TRG si basano la prima sulla non località e la seconda sulla separabilita'. Poiché entrambi i sistemi sono utilizzati nel lavoro dai fisici e' evidente la necessita' di coniugarli in qualche modo. Il dibattito "Einstein vs Bohr" e' stimolante non solo perche' ci fa assistere allo scontro intellettuale tra due padri della fisica moderna; ma anche perche' riassume i fondamenti delle principali visioni interpretative della natura. E' auspicabile che dette diverse posizioni possano fondersi e migliorarsi in una formulazione che ci aiuti a capire maggiormente i misteri della Natura. Ma in quale modo ?

BIBLIOGRAFIA

"Fisica Superiore" di F. de Notaristefani

"La Sintesi einsteiniana" di M.. Born

"Il Paradosso della Realtà Fisica" di G. Tarozzi

"La Luna di Einstein" di D. Lindley

"La Causalità Impossibile" di F. Selleri

PRINCIPIO DI INDETERMINAZIONE

Nel 1927 il fisico tedesco **Werner Heisenberg** scoprì che la natura probabilistica delle leggi della MQ poneva grossi limiti al nostro grado di conoscenza di un sistema atomico. Normalmente ci si aspetta che lo stato di una microparticella in movimento (consideriamo ad esempio un elettrone in rotazione attorno al nucleo) sia caratterizzato completamente ricorrendo a due parametri : velocità e posizione.

Heisenberg postulò invece, che a un certo livello queste quantità sarebbero dovute rimanere **sempre indefinite**. Tale limitazione prese il nome di Principio di Indeterminazione. Questo principio afferma che **maggiore è l'accuratezza nel determinare la posizione di una particella, minore è la precisione con la quale si può accertarne la velocità e viceversa (*)**.

Quando si pensa all'apparecchiatura necessaria per eseguire le misurazioni, questa indeterminazione risulta intuitiva. I dispositivi di rilevazione sono così grandi che la misurazione di un parametro come la posizione è destinato a modificare la velocità. Occorre sottolineare però che le limitazioni in parola, non derivano solo dall'invasiva interazione del mondo macroscopico sul mondo microscopico, **ma sono proprietà intrinseche (ontologiche) della materia. In nessun senso si può ritenere che una microparticella possieda in un dato istante una posizione e una velocità.**

(*) Più precisamente nella misura simultanea delle coordinate x e della quantità di moto p di una particella è **impossibile** ottenere valori x' e p' con indeterminazione piccola a piacere. Infatti, se Δx e Δp denotano rispettivamente l'indeterminazione in x e p , deve esistere la relazione : $\Delta x \cdot \Delta p \geq \frac{h}{4\pi}$ [$h = h/4\pi$]

In modo analogo, in base al principio di indeterminazione è impossibile mediante una osservazione che duri un tempo Δt determinare l'energia di una particella con un'incertezza inferiore a ΔE , legata a Δt dalla relazione : $\Delta E \cdot \Delta t \geq \frac{h}{4\pi}$

Quest'ultima relazione dimostra che in nessun senso si può ritenere che una microparticella possieda in un dato istante una energia definita (Rif. indeterminazione Energia-Tempo).

Si noti infine che il Principio di Indeterminazione è valido per qualsiasi "oggetto", ma in pratica ha conseguenze importanti solo se applicato a particelle di dimensioni atomiche o subatomiche, perché quando si tratta di corpi ordinari, data la piccolezza della costante h esso perde gran parte del suo significato.

L'indeterminazione dal punto di vista sperimentale

Per misurare la POSIZIONE di un oggetto microscopico come un elettrone occorre investirlo con un raggio di luce (fotoni) o comunque qualcosa che in ultima analisi risulta avere all'incirca le **medesime dimensioni dell'elettrone**. Questo fa sì che l'elettrone **risulti perturbato** da questa interazione che ne modifica inesorabilmente la velocità.

La stessa cosa, ma in situazioni opposte, avviene nel caso in cui si voglia conoscere la VELOCITA' di un elettrone ...

Secondo Heisenberg l'effetto dell'interazione tra il raggio di luce incidente e la particella "osservata" può anche essere reso uguale a zero, in tal caso però non si compie alcuna misurazione e non si acquista alcuna conoscenza sulle proprietà del sistema osservato. Se l'interazione è invece diversa da zero, **essa non può essere resa arbitrariamente piccola e non è quindi neppure concettualmente eliminabile**, ma deve avere la **costante di Planck come valore minimo**. Questo perché ogni interazione fisica tra strumento misuratore e microoggetto misurato implica sempre uno scambio di energia per un certo intervallo di tempo, oppure la

cessione di una certa quantità di moto su una certa distanza spaziale. Ora le dimensioni fisiche di "energia per tempo" e di "quantità di moto per spazio" sono proprio quelle dell'**azione**, sono cioè quelle della costante (h) di Planck. Il termine **interazione** usato per ogni processo quantistico di misurazione è quindi vero letteralmente, dato che si ha proprio a che fare con lo scambio della grandezza fisica azione fra lo strumento e il sistema fisico misurato, che a causa delle dimensioni atomiche di quest'ultimo, comporta su di esso una inevitabile perturbazione ...

Un esempio di indeterminazione sperimentale può essere il tentativo di determinazione simultanea della TRAIETTORIA e della POSIZIONE di un oggetto in movimento :

Si supponga di voler determinare, per mezzo di una fotografia, la POSIZIONE di un oggetto in movimento : per esempio una palla di cannone. Naturalmente se la palla percorre la propria parabola a notevole velocità, la fotografia risulterà mossa a meno che si usi un otturatore ad alta velocità.

Più l'oggetto risulterà "fermato", più velocemente sarà scattato l'otturatore. Ma in questo caso si sarà pagato un prezzo nella definizione della traiettoria; risultando la palla di cannone ferma.

D'altro canto usando un otturatore a bassa velocità, si fotograferà una linea indistinta che rappresenta fedelmente la TRAIETTORIA della palla, ma che non dà nessuna indicazione sulla posizione.

In conclusione più si tenta di definire la traiettoria dell'oggetto più si perde informazione sulla posizione e viceversa.

L'indeterminazione dal punto di vista teorico

Come dianzi specificato il fatto che non si riesca a misurare contemporaneamente la velocità e la posizione di una particella non è dovuto soltanto a restrizioni di ordine pratico, ma è un limite obiettivo della natura.

In altri termini la particella allo "stato naturale" NON HA OGGETTIVAMENTE (ontologicamente) una velocità e una posizione.

Questo nasce essenzialmente dalla natura ondulatorio-corpuscolare della materia. **NASCE DAL FATTO DI DOVER METTERE INSIEME ONDE E CORPUSCOLI PER LA RAPPRESENTAZIONE DI UNA STESSA ENTITA'.**

I fenomeni ondulatori :

Un'onda è un ente che vibra nello spazio e nel tempo. Possiamo osservarla in un dato istante di

tempo e allora vediamo una figura periodica nello spazio come nell'esempio di figura A.

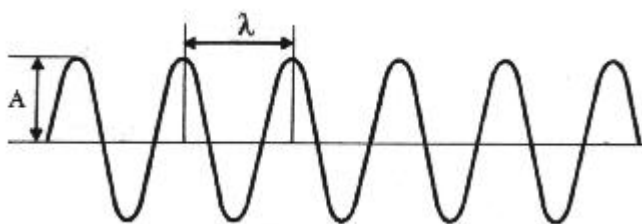


Figura A

Questa forma d'onda è caratterizzata da un'ampiezza A (l'estensione della vibrazione), da una lunghezza d'onda λ (la distanza tra due creste) e da una frequenza n , cioè dal numero di oscillazioni (per coprire la lunghezza λ) compiute nell'unità di tempo.

Come precedentemente visto, de Broglie ha stabilito che ogni grano di energia, luce o materia, possiede una natura ondulatoria. L'informazione sullo stato di moto di una particella è contenuta nella lunghezza d'onda e nella frequenza. La lunghezza d'onda è inversamente proporzionale alla quantità di moto della particella ($\lambda = h/p$). [il che significa che un'onda con piccola lunghezza d'onda corrisponde a una particella che si muove con una grande quantità di moto (e quindi con elevata velocità)]. La quantità di moto p , sarà ovviamente, uguale a h/λ . La frequenza dell'onda sarà infine proporzionale all'energia della particella. Un'onda con frequenza elevata indica che la particella ha grande energia (infatti $E = h n$).

Schrödinger, con la sua celebre equazione, ha stabilito che una particella può occupare tutte le *possibili* posizioni (le "*possibili posizioni*" di Schrödinger diventano "*probabili posizioni*" per Born ...) all'interno dell'onda associata (o **pacchetto d'onda** ...), che dovrà avere estensione limitata nello spazio.

La rappresentazione visiva di un pacchetto d'onda che ha un inizio e una fine può essere raffigurato come il passaggio di un treno di onde che comincia all'istante t_1 e finisce all'istante t_2 (vedere la Figura B).

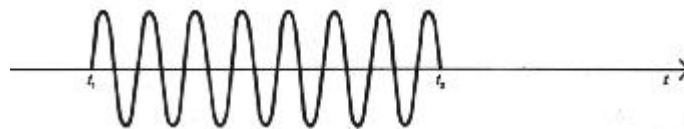


Figura B

L'effetto di questo treno di onde limitato può essere rappresentato mediante una "sinusoide troncata", la cui ampiezza è nulla per $t < t_1$ e per $t > t_2$. Nell'intervallo t_1 a t_2 l'ampiezza non è d'altronde necessariamente come nella figura B anzi, sia la meccanica ondulatoria sia la meccanica quantistica concordano che essa dovrà presentare un massimo al centro t_0 dell'intervallo t_1 a t_2 e un decremento progressivo verso le estremità t_1 e t_2 (vedere la Figura C).

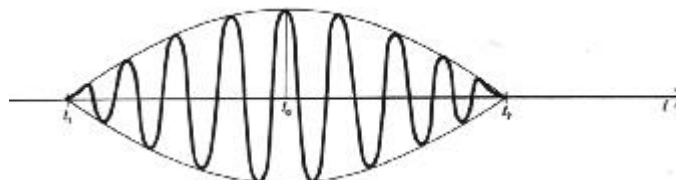


Figura C

All'inizio del XIX secolo il matematico francese Fourier dimostrò che nell'uno e nell'altro caso si può rappresentare analiticamente tale sinusoide troncata come **un**

insieme di sinusoidi indefinite, la cui pulsazione ω varia in maniera continua su un intervallo $D\omega$ intorno a ω_0 . Questo intervallo di pulsazione $D\omega$ è legato all'intervallo $Dt = t_2 - t_1$ dalla relazione : $D\omega \cdot Dt = 2\pi$

Fourier ha dimostrato che è sempre possibile scegliere le ampiezze e le fasi di queste onde componenti in modo tale che la loro sovrapposizione dia un'ampiezza risultante nulla al di fuori dell'intervallo $t_1 t_2$ -per interferenza distruttiva- e che all'interno di questo intervallo l'ampiezza risultante vari nel modo desiderato -per interferenza costruttiva- (vedere la Figura D).

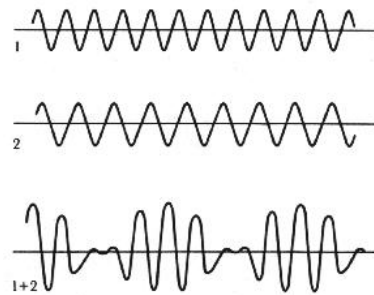


Figura D

Interferenza di onde. Quando si hanno due sorgenti [diverse] di onde elettromagnetiche, ad esempio due sorgenti luminose, l'intensità luminosa misurata in un dato punto non è necessariamente uguale alla somma delle intensità provenienti dalle due sorgenti, ma può essere maggiore o minore. Questo fenomeno può essere spiegato facilmente con l'interferenza delle onde emesse dalle due sorgenti : nei punti in cui si sovrappongono due creste avremo un'intensità luminosa maggiore della somma delle singole intensità. Laddove si sovrappongono una cresta e un ventre avremo un'intensità minore. Il valore esatto dell'interferenza può essere calcolato facilmente. Fenomeni di interferenza di questo tipo si possono osservare ogni volta che si ha a che fare con la radiazione elettromagnetica.

La Figura E infine, mette in chiara evidenza come la curva risultante dalle curve 1 e 2 della Figura D sia formata da una banda di onde di lunghezza differente (risultanti da un annullamento reciproco al di fuori della regione di ampiezza Dx e da un rinforzo reciproco nel suo interno).

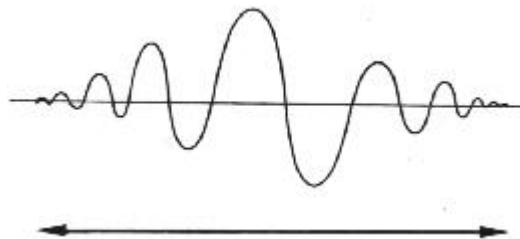


Figura E

D x

Appare chiaro ora che qualora si voglia definire la POSIZIONE di una particella all'interno del pacchetto d'onda (all'interno della curva di Figura E), bisognerà confinare la particella in una "ristrettissima" regione di spazio, la qual cosa seguendo Fourier si ottiene **sovrapponendo insieme molte onde di lunghezza d'onda differente.**

Ma come ci insegna de Broglie lunghezza d'onda è quantità di moto (VELOCITA') sono strettamente collegate e quindi se aumenta il numero degli "enti onda" di differente lunghezza considerati, aumenta anche il numero delle quantità di moto dell'insieme. In definitiva se per definire meglio la posizione di una particella occorre aumentare il numero delle onde (di diversa λ) all'interno del pacchetto d'onda, dovrà aumentare anche il "numero" delle quantità di moto e quindi anche della velocità del pacchetto d'onda ... [$\lambda = h/p$ $\lambda = h/p$ $\lambda = h/p$ $\lambda = h/p$ $\lambda = h/p$...]

Dal confronto delle curve di Figura F, si vede chiaramente che per confinare in un volume più piccolo la particella, occorre comprimere il pacchetto d'onda (curva inferiore). Tale "compressione" può avvenire soltanto a spese di un aumento nel numero delle onde di banda diversa componenti il pacchetto d'onda e quindi delle quantità di moto e quindi della velocità.

Più si tenta di definire la POSIZIONE della particella (confinandola in un volume sempre più piccolo), più questa aumenta la sua VELOCITA' e viceversa.

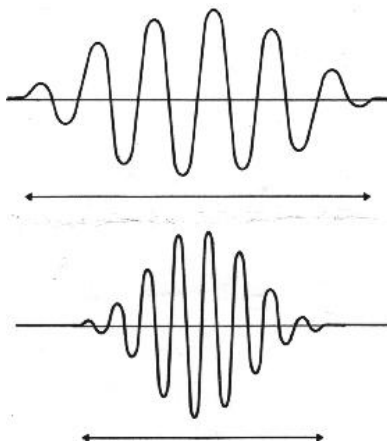
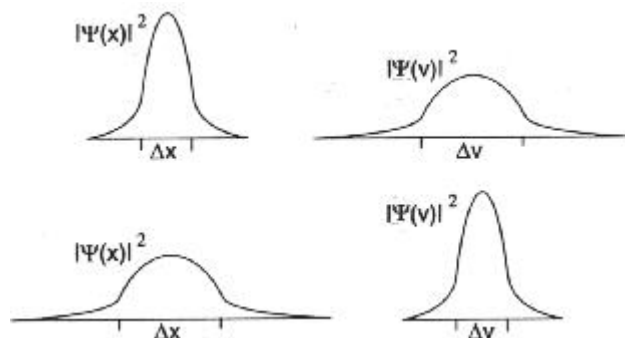


Figura F

La prossima figura mostra come dalle relazioni che legano le funzioni $Y(x)$ e $Y(v)$ segua che allorché una di esse risulti ben concentrata (e quindi la relativa variabile D risulta confinata in un intervallo stretto di valori) l'altra risulti allargata, e quindi la conseguente variabile risulta apprezzabilmente indeterminata. L'argomento rappresenta il modo matematicamente preciso di concludere che le variabili

incompatibili posizione e velocità soddisfano inevitabilmente il Principio di Indeterminazione.



Nota : Occorre fare molta attenzione infine, quando in meccanica quantistica si afferma : "la misurazione modifica la cosa misurata", perché questa asserzione sembra implicare che prima della misura un oggetto quantistico si trovi in un qualche stato definito (ma ignoto), che è stato poi disturbato da un atto di misura. E' vero invece che la misurazione conferisce una DETERMINAZIONE ad una quantità che in precedenza era (oggettivamente) indefinita ; non si può assegnare alcun significato a una quantità finché non la si misura.

L'INDETERMINAZIONE ENERGIA/TEMPO

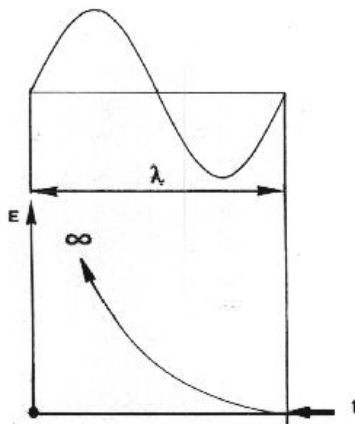
- Si supponga di voler determinare l'energia di un fotone. Secondo il formalismo ideato da Planck, l'energia di un fotone è direttamente proporzionale alla frequenza della luce ($E = h \nu$) : se si raddoppia la frequenza, anche l'energia diventa doppia. Un sistema pratico per misurare l'energia è quindi quello di misurare la frequenza dell'onda luminosa, il che si può fare contando il numero di oscillazioni (il susseguirsi cioè di massimi -le creste- e di minimi -gli avvallamenti-) in un dato intervallo di tempo. Per poter applicare questa procedura occorre comunque che si verifichi almeno una e preferibilmente, più di una oscillazione completa, che richiede un intervallo di tempo definito. L'onda deve passare da un massimo ad un minimo, e poi di nuovo tornare a un massimo. Misurare la frequenza della luce *in un tempo inferiore* a quello occorrente per un'oscillazione completa è evidentemente impossibile, anche in via di principio. Per la luce visibile il tempo occorrente è ridottissimo (un milionesimo di milionesimo di secondo). Onde elettromagnetiche con lunghezze d'onda maggiore e frequenza minore, come le onde radio, possono richiedere qualche millesimo di secondo per compiere un'oscillazione completa. I fotoni delle radioonde hanno, in confronto, un'energia molto bassa; invece i raggi gamma

oscillano migliaia di volte più rapidamente della luce e le energie dei fotoni sono maggiori di migliaia di volte.

Queste semplici considerazioni mettono in evidenza l'esistenza di un limite fondamentale alla precisione con cui è possibile misurare la frequenza di un fotone (e quindi la sua energia) in un dato intervallo di tempo. ***Se l'intervallo è inferiore a un periodo dell'onda, l'energia è quanto mai indeterminata*** ; esiste quindi una relazione di indeterminazione che lega l'energia al tempo, identica a quella già vista per posizione e quantità di moto. Per avere una determinazione esatta dell'energia si deve effettuare una misurazione "relativamente lunga", ma, se ciò che ci interessa è invece l'istante in cui si verifica un evento, lo si potrà determinare in modo esatto solo a spese dell'informazione sull'energia. Ci si trova così a dover scegliere tra l'informazione sull'energia e l'informazione sul tempo, che presentano un'incompatibilità analoga a quella per posizione e moto. Questa nuova relazione di indeterminazione ha conseguenze molto importanti ...

Va (ri)sottolineato che i limiti alle misurazioni sia di energia e tempo, sia di posizione e quantità di moto, non sono semplici insufficienze tecnologiche, ma proprietà assolute e intrinseche dalla natura. **In nessun senso si può pensare che un fotone (o un elettrone, o ecc.) possieda realmente un'energia ben definita in un dato istante.** Per i fotoni energia e tempo sono due caratteristiche incompatibili; quale delle due si manifesti con maggior precisione dipende solo dalla natura della misurazione (osservazione) che **si sceglie** di effettuare ...

$\Delta E \cdot \Delta t \geq h$ [$h = h/4\pi$] da ciò deriva che $\Delta E = h/\Delta t$ Da questa formula si vede che più Δt diventa piccolo (cioè più tende a 0) più ΔE diventa grande (cioè tende all'infinito).



Appunti integrati su

MODELLI QUANTISTICI DELLA MEMORIA

E.Pessa, A.Ferretti

- Nel 1967 **L.M.Ricciardi e H.Umezawa** proposero il primo modello quantistico di memoria
- Nel 1968 **Frohlich** ipotizzò l'esistenza di vibrazioni elettriche stabili, coerenti, ad alta frequenza di origine quantistica nei sistemi biologici
- Nel 1970 **Walker** avanzò la tesi secondo la quale i fenomeni coscienti possono essere spiegati solo in termini quantistici
- Molti scienziati, tra i più noti **K.Pribram** e **R.Penrose**, hanno scritto sull'argomento
- Lo studio della psicologia deve rispondere alle seguenti domande:
 - 1) Perché le idee della QM, applicate allo studio dei fenomeni fisici, dovrebbero essere rilevanti per lo studio dei fenomeni psichici?
 - 2) I risultati perseguibili nello studio dei fenomeni psichici con l'ausilio delle idee della QM, non potrebbero essere raggiunti usando i metodi convenzionali delle scienze psicologiche?
 - 3) Ammesso che le idee della QM siano rilevanti e necessari, sarebbero anche sufficienti?
- I concetti della QM possano essere usati per costruire un modello di memoria, in particolare quello proposto da G.Vitiello (1995)
- Uno dei maggiori problemi in psicologia è quello del collegamento tra due possibili livelli di descrizione, quello comportamentale (o macroscopico) e quello dei processi mentali (o microscopico). In ambito comportamentista il confronto fu tra approccio **micromolare** (o molecolare) e **molare**. Rappresentanti di questi due paradigmi furono rispettivamente Guthrie e Skinner. In ambito cognitivista il confronto fu tra l'approccio computazionale simbolico e l'approccio subsimbolico. Questa dicotomia tocca anche la coscienza, intesa a partire da W.James come un processo che accompagna la selezione di stimoli e azioni. Il problema è evidente anche nello studio di funzioni specifiche, come la percezione visiva, che ha un aspetto gestaltico e dipendente dal contesto, come pure fattori locali.
- I modelli teorici psicologici ricadono in una delle seguenti categorie:
 - a) Modelli basati sulla teoria dell'informazione, sulla teoria del calcolo simbolico, sulla teoria della probabilità (p.es. HIP)
 - b) Modelli connessionistici basati sulle reti neurali, modelli di memoria associativa, reti semantiche, modelli comportamentisti di apprendimento
 - c) Modelli basati sul concetto di campo, preso a prestito dalla fisica classica, come le teoria gestaltiche della percezione visiva
- La b) ha il vantaggio di poter dare descrizioni microscopiche. Tuttavia è carente nella capacità di estrarre descrizioni macroscopiche, sostanzialmente per due motivi:

1) Non c'è una misura per definire lo stato macroscopico della rete, che abbia un ruolo analogo a quello dell'energia di un sistema fisico

2) Le unità che formano la rete sono considerate distinguibili, sia per la posizione che occupano che per il ruolo che giocano nel rappresentare l'informazione

- Nella QT queste due condizioni non si verificano. La QT è l'unica che ammetta una descrizione microscopica e macroscopica.
- Il termine **QT** si riferisce a due diversi sistemi teorici, la **meccanica quantistica (QM)**, elaborata da Schrödinger, Heisenberg, Born e Jordan, e la **teoria dei campi quantistici (QFT)**, sviluppata da Dirac, Feynmann, Schwinger, Dyson.
- La QM descrive sistemi di particelle con un numero finito di variabili, o gradi di libertà. Una delle sue conseguenze più importanti è il principio di indeterminazione
- La QFT è molto più complessa, e descrive sistemi con infiniti gradi di libertà, come i campi, dove c'è una variabile dipendente in ogni punto dello spazio. A differenza della QM, la QFT è capace di ammettere sin dal principio due livelli di descrizione: microscopico e macroscopico. La QFT non può essere considerata ancora una teoria completa, e richiede ulteriori estensioni e generalizzazioni per poter essere applicata allo studio dei fenomeni emergenti nei sistemi più complessi di quelli tradizionalmente considerati della fisica (Umezawa, 1993).
- Tra i fenomeni quantistici macroscopici ricordiamo la superconduttività, la superfluidità, l'effetto laser.
- La capacità di previsione della QFT si basa su due meccanismi:
 - 1) L'uso di formalismi di descrizione dinamiche microscopiche analoghi a quelli della meccanica classica
 - 2) L'ipotesi di indistinguibilità delle unità microscopiche

La Meccanica Statistica dimostra che se la 2) non vale gli unici fenomeni macroscopici osservabili sono fluttuazioni temporanee attorno a uno stato di equilibrio globale (Rumer e Ryvkin, 1980). Questo significa che le reti neurali così come sono concepite normalmente non sono veri modelli microscopici, ma modelli macroscopici camuffati.

- La possibilità di considerare la materia vivente come una delle fasi della materia ordinaria è una delle prospettive più interessanti della fisica contemporanea.
- Il comportamento della materia vivente, caratterizzato da riduzione di entropia e da reazioni chimiche ben ordinate nello spazio e ben sequenziate nel tempo, sfugge a una trattazione fisica classica
- Il primo modello quantistico del funzionamento cervello apparve nel 1967, in un articolo di **L.M. Ricciardi** e **H. Umezawa** apparso sulla rivista *Kybernetik*. Gli autori si proponevano di rispondere a una serie di domande sul "modo in cui il cervello elabora l'informazione ricevuta dall'esterno". Essi riconoscevano i limiti dei modelli matematici usati fino ad allora, dove i neuroni erano rappresentati per mezzo di un'equazione algebrica e le connessioni sinaptiche da coefficienti.

Per Umezawa e Ricciardi non è necessario conoscere l'evoluzione temporale del comportamento di ogni singolo neurone per capire il comportamento del cervello nel suo insieme: il funzionamento stesso del cervello non dipende in maniera critica dalla perfetta efficienza di ogni suo componente elementare. Gli autori ipotizzarono anche l'esistenza di nuovi complessi processi nascosti dietro l'attività delle cellule cerebrali, ciò per spiegare i risultati sperimentali di **John** (1967) il quale aveva mostrato come, a seguito di particolari stimolazioni, risposte correlate e quasi simultanee potevano insorgere in regioni cerebrali anche lontane tra loro. Si affacciava l'idea di un effetto di natura quantistica. La resistenza delle tracce mnestiche a interventi di ablazione e trattamenti farmacologici faceva ritenere che le tracce dovevano essere diffuse nell'intero cervello, e non cablate in circuiti ben definiti. La coscienza non può richiamare più informazioni simultaneamente, sicché il processo di richiamo è di tipo seriale e non parallelo. Il modello considera il cervello come un *sistema* che, interagendo con gli *stimoli* ambientali, cambia il suo *stato*. Gli stati *stazionari* o *quasi-stazionari* del cervello isolato sono descritti da opportune variabili dinamiche. In termini quantistici gli stati sono classificati per mezzo dei numeri quantici associati all'invarianza del sistema sotto certe trasformazioni. Gli stimoli, codificati nel cervello, perdurano anche dopo la loro cessazione. La specificazione del numero quantico sarebbe equivalente al codice stesso. Solo quelle variabili invarianti rispetto a una opportuna trasformazione T sono responsabili dell'immagazzinamento dell'informazione caratterizzata dal codice corrispondente. Queste variabili potevano essere classificate in insiemi indipendenti, labellati con i numeri quantici, che si trasformano indipendentemente sotto le suddette trasformazioni. Lo stimolo cambia i valori delle variabili producendo stati eccitati dei numeri quantici associati. Questa prima trasformazione poteva essere considerata come una MBT. Il codice, l'informazione contenuta negli stati eccitati del sistema, poteva poi essere trasferito nello stato di energia minima del sistema, per condensazione nello stato fondamentale. Questo processo poteva essere considerato una sorta di apprendimento. Nella **condensazione di Bose-Einstein**, la condensazione dei bosoni aventi uno specifico numero quantico trasferisce nello stato di minima energia il numero quantico stesso (allo stesso modo in cui la condensazione di un'onda di spin crea uno stato polarizzato stabile nel caso del ferromagnetismo). A tal proposito osserviamo che:

- a) La condensazione di Bose-Einstein si verifica solo se le particelle sono microscopicamente indistinguibili
 - b) La condensazione avviene nello spazio dei momenti, cioè nello stato condensato tutte le particelle hanno la stessa velocità
 - c) Nello stato condensato, esistono interazioni a lungo raggio tra le particelle che producono forti correlazioni nella velocità di particelle anche molto distanti
- Il comportamento del cervello potrebbe essere caratterizzato in termini di variabili che obbediscono alla statistica di Bose-Einstein. Il funzionamento del cervello

sarebbe un effetto della rottura spontanea delle sue diverse simmetrie dinamiche. Mentre la MLT sarebbe associata a uno stato stabile di minima energia, la MBT sarebbe associata a stati metastabili eccitati.

- Nei modelli classici queste variabili possono essere identificate con l'attività dei neuroni della rete, ma a Ricciardi e Umezawa era chiaro che le usuali unità funzionali della neurofisiologia classica (neuroni e sinapsi) non potevano essere usate per descrivere un sistema quantistico.
- Successivamente **Stuart, Takahashi e Umezawa** (1978-79) identificarono questi oggetti nei **corticoni** (*corticons*) e nei **simmetroni** (*symmetrons*). Il corticone è un modo collettivo associato al cervello prima che questo riceva l'informazione esterna da memorizzare. Il simmetrone è il trasporto delle correlazioni a lungo raggio dopo l'ingresso dell'informazione esterna, che dà luogo alla rottura spontanea di simmetria nel sistema corticon.
- In questi articoli venivano messi in luce anche i limiti dei modelli quantistici, come il **problema della capacità di memoria**: l'arrivo di una nuova informazione doveva portare il sistema da uno stato di minima energia a un altro stato di minima energia, che distruggeva così l'informazione precedente.
- Questo problema è stato risolto da **G. Vitiello** (1995), osservando che noi possiamo ricordare solo il passato, sicché il processo di memorizzazione produce anche una rottura della simmetria del time-reversal del sistema., cioè una perdita di invarianza rispetto alla inversione dell'asse dei tempi. In altre parole, l'evoluzione del sistema cambia se noi cambiamo la direzione del tempo (una sorta di irreversibilità). Il sistema è quindi dissipativo, e come tale non può essere trattato coi metodi della fisica classica, che è basata sul principio di conservazione dell'energia, non soddisfatto dai sistemi dissipativi. Vitiello ha superato questa difficoltà osservando che un sistema dissipativo è un sistema aperto che cede energia all'esterno. Ma allora il sistema globale comprendente il sistema dissipativo e l'ambiente soddisfa ancora al principio di conservazione dell'energia. Possiamo modellare l'ambiente come un sistema identico al sistema dissipativo in esame, ma con inversione del tempo, una sorta di specchio caratterizzato da modi collettivi "specchiati" (si ricordi che invertendo l'asse dei tempi l'energia dissipata diventa energia assorbita). Questo fatto introduce una perfetta ortogonalità tra le varie informazioni immagazziate, di modo che le tracce sono indipendenti tra di loro.
- Il modello di Vitiello è stato implementato su una rete di Hopfield modificata:
 - a) le unità sono lineari
 - b) le unità possono avere autoconnessioni
 - c) ogni unità è spazialmente localizzata in un reticolo 2D (M righe, N colonne)
 - d) le connessioni sono solo locali: l'unità in posizione (i,j) è connessa solo alle unità $(i,j-1)$, $(i,j+1)$, $(i-1,j)$, $(i+1,j)$, diciamo il dominio $D(i,j)$. Una particolare condizione vale sulle unità al bordo.

- e) Non c'è differenza tra fase di apprendimento e fase di richiamo: la rete apprende continuamente.
- f) La funzione di attivazione è

$$x_{jk}(t+1) = KP_{jk}(t)$$

$$P_{jk}(t) = h + \alpha x_{jk}(t) - \gamma [x_{jk}(t)]^2 + I_{jk}(t) + \delta$$

$$h = \sum_{r,s \in D(j,k)} w_{jkrs} x_{rs}(t)$$

$$\delta = \sum_{r,s \in D(j,k)} [\alpha_{MAX} - x_{rs}(t)] |x_{jk}(t)|$$

dove $I_{jk}(t)$ è l'input per l'unità di coordinate (j,k) .

- Il comportamento dissipativo della rete è determinato dai termini con segno negativo. α_{MAX} serve a controllare possibili condizioni di overflow.
- La legge di apprendimento è data dalla regola di Hebb (con ritardo sinaptico 1):

$$\Delta w_{jkrs} = x_{jk}(t) x_{rs}(t-1)$$

- L'evoluzione libera di una rete con 50×50 unità è stata osservata per $t = 0, \dots, 200$ passi, dopo aver applicato valori di x_{ij} scelti a caso in $[-1, 1]$, ma nulli ai bordi, e valori w_{jkrs} scelti a caso in $[-0.1, 0.1]$ ($K=1$, $\alpha = 0.1$, $\alpha_{MAX}=1$, $\gamma = 0.295$). Si è osservato una distribuzione delle attivazioni sinusoidale nello spazio e nel tempo, smorzata nel tempo.
- Per verificare se è possibile interpretare queste distribuzioni come modi collettivi "wave-like" si è applicato uno stato iniziale del tipo

$$x_{ij}(0) = A \sin(k_x i) \sin(k_y j) \quad k_x = 2\pi n/N, \quad k_y = 2\pi n/M$$

dove n è un intero.

- L'analisi di Fourier nel tempo delle attività di un certo numero di unità ha mostrato modi collettivi variabili a seconda dello stato iniziale scelto. Possiamo pensare che la rete abbia molti stati fondamentali (*ground states*).
- Per immagazzinare informazione nella rete si sono usati vettori binari \mathbf{p} con un numero pari V di componenti, dove le prime $V/2$ componenti, $p_1, \dots, p_{V/2}$, codificano un numero naturale n_1 , mentre $p_{V/2+1}, \dots, p_V$ codificano per un numero naturale n_2 :

$$I_{jk} = A \sin(2\pi n_1/M) \sin(2\pi n_2/N) \quad (1)$$

- Distinguiamo allora due fasi

- a) **Fase di memorizzazione.** L'input costante I_{jk} è applicato per un certo numero di passi, in cui i coefficienti della rete si modificano in base alla regola di Hebb
 - b) **Fase di evoluzione libera.** Evoluzione libera della rete verso lo stato fondamentale, con coefficienti di connessione fissi.
- La fase di evoluzione libera deve risentire in qualche modo dell'informazione memorizzata. Dato che questo effetto tende a svanire rapidamente, a causa della natura dissipativa della rete (i valori di attivazione $x_{ij}(t)$ diventano subito troppo piccoli per essere usati), occorre definire l'ampiezza minima Δt della finestra temporale, a partire dall'inizio della fase di evoluzione libera, che contenga in sé tutte le informazioni relative al nuovo stato fondamentale della rete. In base alle idee di Vitiello la dissipazione è causata da un'entità (l'ambiente esterno) che fornisce un'energia ΔE alla rete. In base alla QT, ΔE e Δt sono collegate dal **principio di indeterminazione di Heisenberg**

$$\Delta E \cdot \Delta t \geq h/4\pi$$

da cui possiamo stimare

$$\Delta t \approx h/\Delta E$$

Per calcolare ΔE possiamo eguagliarlo al lavoro fatto dalle forze dissipative

$$\Delta E = \int_{x(0), x(t)} F dx$$

Le forze dissipative F sono rappresentate dai termini

$$\Delta x_{jk}(t) \propto -|x_{jk}(t)| \sum_{r,s \in D(j,k)} x_{rs}(t) - \gamma [x_{jk}(t)]^2$$

Passando al continuo, posto $x (|x_{jk}(t)|)$ e $y (\sum_{r,s \in D(j,k)} x_{rs}(t))$,

$$dx/dt \propto -xy - \gamma x^2$$

Derivando ancora e riusando la precedente abbiamo

$$d^2x/dt^2 \propto xy^2 + 3\gamma x^2y + 2\gamma^2 x^3$$

Possiamo interpretare la precedente come l'“accelerazione” di una massa unitaria prodotta dalla forza F , avente la stessa espressione. Sostituendo nell'integrale e trascurando i termini con γ^2 , e usando il fatto che x è sempre molto vicino a y , abbiamo

$$\Delta E = (\gamma + 1/2) |x^4 - x_0^4| \approx (\gamma + 1/2) x_0^4$$

Quindi

$$\Delta t \approx h / [(\gamma + 1/2) x_0^4]$$

Dato che il sistema è a tempo discreto (il passo temporale è 1) $h=1$, sicché

$$\Delta t \approx 1 / [(\gamma + 1/2) x_0^4]$$

Con un valore tipico $x_0 = 0.7$, $\gamma = 0.295$, si ha $\Delta t \approx 5$. Questo significa che l'effetto dell'immagazzinamento dell'informazione nella rete è misurabile già a partire dai primi 5 passi dell'evoluzione libera seguente.

- Si è voluto sottoporre la rete a una prova di riconoscimento. Questo tipo di test consiste nel far apprendere al soggetto una lista di items (*lista di studio*). Successivamente si presenta una seconda lista di uguale lunghezza (*lista di test*), per ciascun item della quale il soggetto deve dire se apparteneva alla lista di studio. Nei soggetti umani si osserva l'*effetto lunghezza della lista (list-length effect)*, cioè una riduzione delle prestazioni all'aumentare del numero di items nelle liste (R). L'obiettivo è verificare se anche la rete neuronale descritta presenti questa caratteristica.
- In una prima prova si è sottoposta una lista di studio di $R = 20$ vettori di 10 componenti scelte a caso (gli items, \mathbf{p}^s con $s = 1, \dots, 20$). Ogni vettore, codificato tramite la (1), è stato applicato per 10 passi di tempo, e a seguire 8 passi di evoluzione libera, in cui si sono registrate le attività, $x_{ij}^s(t)$. Nella fase di riconoscimento i coefficienti di connessione sono stati bloccati. Ciascun vettore della lista test (\mathbf{t}^q con $q = 1, \dots, 20$) è stato applicato per 1 passo, e a seguire si sono registrate le attività per i successivi 5 passi, $x_{ij}^q(t)$. Poi si sono calcolati i coefficienti di correlazione di Bravais-Pearson $C^{s,q}(t)$ e la loro media entro la finestra temporale Δt , $AC^{s,q}(t)$.
- In una seconda prova si è ripetuto il tutto con liste di studio e di test di $R=10$ items.
- Si è osservato che con $R=10$ items $AC^{s,q}(t) > 0.2$ iff l'item della lista di test era presente nella lista di studio, altrimenti $AC^{s,q}(t) \approx 3 \times 10^{-5}$. Se assumiamo 0.2 come soglia di classificazione abbiamo un riconoscimento perfetto su tutta la lista, com'è ovvio. La situazione cambia molto con liste di $R=20$ items. In tal caso è sempre $AC^{s,q}(t) < 0.2$, cioè non vi è nessun riconoscimento (anche se in realtà 6 dei 20 item della lista di test erano effettivamente presenti nella lista di studio). Tutto questo sembra indicare la presenza di un *list-size effect*. Una riprova è che per $R=1$ si è trovato $AC^{s,q}(t) = 0.94$ (stavolta mediando su $\Delta t = 64$ passi).

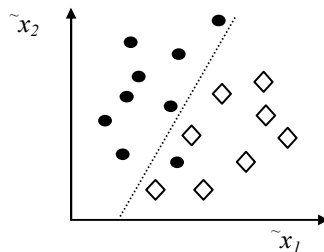
Appunti integrati su
NEURAL NETWORKS FOR PATTERN RECOGNITION
C.M.Bishop

CAP 1. PATTERN RECOGNITION STATISTICO

- Consideriamo il problema di associare un'assegnazione del vettore

$$\mathbf{x} = (x_1, \dots, x_d)^T \quad 0 \leq x_i \leq 1$$

- a una delle due classi C_1 (l'immagine di una "a") e C_2 (l'immagine di una "b").
- *Def.* Un **data set** (o un **campione**) è un insieme di vettori per i quali già esiste l'associazione con una delle due classi.
- Per immagini 256x256, ogni vettore è un punto in uno spazio a $d = 16536$ dimensioni. Ciascun asse codifica il livello di grigio di un determinato pixel, supponiamo con una risoluzione di 8 bit. E' assurdo pensare di memorizzare tutte le immagini possibili ($2^{256 \cdot 256 \cdot 8}$).
- Una tecnica per aggirare il **problema della dimensionalità dell'input**, cioè dell'elevato numero di variabili di ingresso, è quella combinare molte di queste variabili in un piccolo numero di nuove variabili (**feature extraction**). Nel nostro esempio una di queste nuove variabili potrebbe essere il rapporto tra altezza e larghezza del carattere, \tilde{x}_1 .
- La distribuzione di \tilde{x}_1 sugli elementi del *data set* associati a C_1 avrà una media inferiore rispetto a quella della distribuzione di \tilde{x}_1 sugli elementi del *data set* associati a C_2 . Le due distribuzioni potranno sovrapporsi in un certo intervallo, sicché il parametro \tilde{x}_1 non permette una distinguibilità assoluta. Una possibilità è quella di fissare un valore di soglia, proprio nel punto di intersezione delle due distribuzioni.
- L'aggiunta di un altro parametro \tilde{x}_2 riduce ulteriormente la possibilità di errori di classificazione.



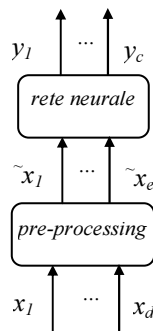
- Notiamo che con due parametri la separazione tra le due classi (linea tratteggiata) è migliore rispetto a quella che si avrebbe con uno solo dei due (le distribuzioni

dei punti di proiezione su uno dei due assi hanno sempre una maggiore sovrapposizione)

- Aggiungendo sempre più parametri (indipendenti) i risultati migliorano fino a un certo punto.
- Formalmente, il nostro algoritmo di classificazione è una funzione

$$y_k = f(\mathbf{x}, \mathbf{w}) \quad k=1, 2, \dots, c$$

- dove \mathbf{w} è un vettore di parametri, e $y_k = 1 \Leftrightarrow \mathbf{x}$ è classificato in C_k .
- In una rete neurale \mathbf{w} sono i **pesi** (*weights*), e la loro determinazione sulla base del *data set* è detta **addestramento** (*training*) o **apprendimento** (*learning*)
- Sia i problemi di **classificazione** che di **regressione** (dove gli output sono variabili continue) sono casi particolari del problema dell'*approssimazione di funzioni*. Nei problemi di classificazione cerchiamo di approssimare le probabilità di appartenenza delle classi C_k in funzione delle variabili di ingresso \mathbf{x} . Nei problemi di regressione cerchiamo di approssimare la *funzione di regressione*.
- Lo schema di principio di una applicazione di rete neurale è



- dove lo stadio di pre-processing, ed eventualmente di post-processing, sono trasformazioni fisse, mentre la rete neurale è una trasformazione a parametri adattativi.
- *Def.* Un sistema di riconoscimento visivo è **invariante alla traslazione (scala)** \Leftrightarrow opera indipendentemente dalla posizione (dimensione) del pattern nel vettore di ingresso.
- Le proprietà di invarianza sono **informazioni a priori** (*prior knowledge*) che si aggiungono al *data set* di addestramento.
- Il **problema della dimensionalità** (*curse of dimensionality*). Un modo inefficiente di specificare una funzione multivariata non lineare da \mathbf{x} in y sarebbe quello di suddividere il range di ciascuna variabile di ingresso x_i in M intervalli, cioè lo spazio di \mathbf{x} in M^d celle discrete, e specificare il valore di y per ciascuna di esse. Chiaramente questo metodo è impraticabile in quanto il data set richiesto aumenta esponenzialmente con d .
- Metodi più efficienti tengono conto delle correlazioni che esistono tra le variabili di ingresso, da cui il concetto di **dimensionalità intrinseca**. Inoltre y generalmente non varia arbitrariamente per ingressi abbastanza vicini, sicché è sempre possibile

inferire i valori di y per punti intermedi dove non sono disponibili dati, con un procedimento simile all'interpolazione.

- Un'applicazione è il *fitting* di una curva polinomiale:

$$y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \dots + w_M x^M = \sum_{j=0..M} w_j x^j$$

- Il *data set* consiste di N punti (x^n, t^n) . Il procedimento consiste nel determinare \mathbf{w} in modo da minimizzare una qualche funzione d'errore, per esempio

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1..N} [y(x^n, \mathbf{w}) - t^n]^2$$

- Sia $\mathbf{w}^* \mid E(\mathbf{w}^*) = \min\{E(\mathbf{w})\}$
- Notiamo che y è non lineare in x , ma lineare in \mathbf{w} . Funzioni che sono lineari nei parametri adattativi sono dette **modelli lineari**.
- La minimizzazione di E è un caso di **apprendimento supervisionato**, in quanto sono noti i valori *target* $\{t^n\}$. In altri problemi, come la determinazione della distribuzione $P(\mathbf{x})$, in cui non si dispone di $\{t^n\}$, si tratta di un **apprendimento non supervisionato**. Nel caso di **apprendimento rinforzato** (Hertz, 1991) non si fornisce un set di output desiderati $\{t^n\}$, ma si accetta o meno l'output prodotto spontaneamente.
- Dato un **training data set** (x^n, t^n) , e un **test data set** (x^s, t^s) su cui verificare la capacità di **generalizzazione** del modello polinomiale, i risultati migliori si ottengono per M non troppo piccolo né troppo grande. Un numero di gradi di libertà troppo grande aumenta infatti la variabilità del modello $y(x, \mathbf{w}^*)$ (**overfitting**).
- Per verificare la bontà di generalizzazione di $y(x, \mathbf{w}^*)$ a nuovi dati (il *test data set*) conviene usare una espressione dell'errore normalizzata rispetto alla numerosità del set

$$E^{RMS} = \left\{ (1/N) \sum [y(x, \mathbf{w}^*) - t]^2 \right\}$$

- Mentre $E^{RMS}(M)$ decresce monotonamente se valutato sul *training data set*, raggiunge un minimo se valutato sul *test data set*. Un M troppo piccolo produce modelli con alto **bias** e bassa **varianza**, mentre elevati valori di M producono modelli con alta **varianza** e basso **bias**. L'ottimo si ha con un compromesso tra queste due opposte tendenze.
- Riassumendo, la complessità di un modello può essere fatta coincidere col numero di parametri liberi M del modello stesso (l'ordine del polinomio nell'esempio precedente, o il numero di unità nascoste in una rete neurale). A parità di *data set*, la complessità del modello con le migliori capacità di generalizzazione non è né troppo alta né troppo bassa.
- Estendendo il concetto del fitting polinomiale a $d > 1$ ingressi avremo bisogno di d^M parametri liberi. In tal caso per determinare tutti i parametri con sufficiente

precisione occorreranno *data sets* di cardinalità proporzionale. Ma esistono modi alternativi di rappresentare una funzione non lineare multivariata, p.es. come sovrapposizione di funzioni non lineari univariate. Questo è il modello di una rete neurale.

- **Barron** ha calcolato che per una rete neurale a M unità nascoste $E(M) = O(1/M)$, indipendentemente da d . Invece $E(M) = O((1/M)^{2/d})$ nel caso in cui i parametri liberi sono i coefficienti di un qualunque sviluppo in serie di funzioni definite
- Il prezzo da pagare con le reti neurali è il costo computazionale dell'algoritmo di apprendimento, che è non lineare. Una complicazione aggiuntiva è che $E(\mathbf{w})$ ha dei minimi locali.
- Il **Teorema di Bayes** (**Thomas Bayes**, 1702-1761) è alla base dell'approccio statistico al *pattern recognition*. Il problema è quello di minimizzare la probabilità di errore di classificazione di un nuovo pattern \mathbf{x} in una delle classi C_k senza ancora aver operato trasformazioni su \mathbf{x} (o misure di \tilde{x}_l). Supponiamo che \tilde{x}_l , a valori in $\{X^l\}$, sia la caratteristica del pattern \mathbf{x} che è stata usata per classificare gli input del training *data set*. Si ha

$$P(C_k|X^l) = P(X^l|C_k)P(C_k)/P(X^l) = P(X^l|C_k)P(C_k)/\sum_{j=1..c}P(X^l|C_j)P(C_j)$$

- dove $P(C_k|X^l)$ sono le **probabilità a posteriori**, $P(C_k)$ le **probabilità a priori**, e $P(X^l|C_k)$ le probabilità **condizionali di classe** (*class conditional pr.*)
- Oss. $\sum_{j=1..c}P(C_j) = 1$
- Per un pattern \mathbf{x} avente feature X^l la probabilità di classificazione errata è minima se associamo \mathbf{x} alla classe C_k per la quale è massima $P(C_k|X^l)$.
- D'ora in poi chiamiamo \mathbf{x} il vettore di input già processato dallo stadio di pre-processing (quello che prima era la feature \tilde{x}_l).
- Se le x_i sono variabili continue, definiremo la densità di probabilità $p(\mathbf{x})$ in modo tale che

$$P(\mathbf{x} \in \mathcal{R}) = \int_{\mathcal{R}} p(\mathbf{x}) d\mathbf{x}$$

- **Def. Valore atteso** di una funzione $Q(\mathbf{x})$ è

$$\mathcal{E}[Q] = \int Q(\mathbf{x}) p(\mathbf{x}) d\mathbf{x}$$

- Dati N punti $\mathbf{x}^1, \dots, \mathbf{x}^N$, vale l'approssimazione

$$\mathcal{E}[Q] \approx (1/N) \sum_{n=1..N} Q(\mathbf{x}^n)$$

- Il **teorema di Bayes** si riscrive

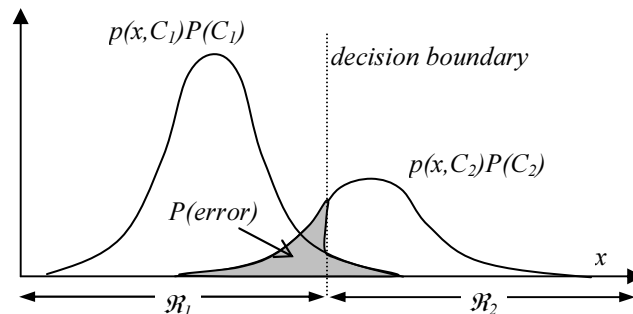
$$P(C_k|\mathbf{x}) = p(\mathbf{x}|C_k)P(C_k)/p(\mathbf{x}) = p(\mathbf{x}|C_k)P(C_k)/\sum_{i=1..c}p(\mathbf{x}|C_i)P(C_i)$$

- La probabilità a posteriori $P(C_k|\mathbf{x})$ ci dà la probabilità che un vettore osservato \mathbf{x} appartenga alla classe C_k . La probabilità di errore è minimizzata scegliendo la classe C_k avente massima probabilità a posteriori, cioè tale che

$$P(C_k|\mathbf{x}) > P(C_j|\mathbf{x}), \quad \forall j \neq k$$

- Un **classificatore di patterns** (*pattern classifier*) è un'applicazione che associa ogni punto nello spazio delle features a una delle c classi. Possiamo immaginare lo spazio delle features suddiviso in c regioni di decisione $\mathcal{R}_1, \dots, \mathcal{R}_c$, tali che un punto che cade nella regione \mathcal{R}_k viene associato alla classe C_k . I bordi di queste regioni sono noti come **bordi di decisione** (*decision boundaries*) o **superfici di decisione** (*decision surfaces*).
- Il punto di decisione ottimo per un problema di classificazione di una variabile su due classi ($c=2$) si ottiene minimizzando la probabilità di errore. Detta $P(x \in \mathcal{R}_u, C_v)$ la probabilità congiunta che x appartenga a \mathcal{R}_u e che sia C_v la vera classe di appartenenza di x , abbiamo

$$\begin{aligned} P(\text{error}) &= P(x \in \mathcal{R}_1, C_2) + P(x \in \mathcal{R}_2, C_1) = \\ &= P(x \in \mathcal{R}_1 | C_2) P(C_2) + P(x \in \mathcal{R}_2 | C_1) P(C_1) = \\ &= \int_{\mathcal{R}_1} p(x|C_2) P(C_2) dx + \int_{\mathcal{R}_2} p(x|C_1) P(C_1) dx \end{aligned}$$



- Le stesse considerazioni si applicano al caso d -dimensionale su c classi.

$$P(\text{correct}) = \sum_{k=1..c} P(\mathbf{x} \in \mathcal{R}_k, C_k) = \sum_{k=1..c} P(\mathbf{x} \in \mathcal{R}_k | C_k) P(C_k) = \sum_{k=1..c} \int_{\mathcal{R}_k} p(\mathbf{x}|C_k) P(C_k) d\mathbf{x}$$

$P(\text{correct})$ è massimizzata scegliendo $\{\mathcal{R}_k\}$ in modo tale che $\mathbf{x} \in \mathcal{R}_k$ quando

$$p(\mathbf{x}|C_k) P(C_k) > p(\mathbf{x}|C_j) P(C_j), \quad \forall j \neq k$$

Inoltre i *decision boundaries* saranno quelle regioni in cui le funzioni discriminanti si eguagliano. P.es. se \mathcal{R}_k e \mathcal{R}_j sono regioni contigue, il *decision boundary* tra di esse sarà

$$\mathbf{x} \mid p(\mathbf{x}|C_j)P(C_j) = p(\mathbf{x}|C_k)P(C_k)$$

- Il problema della classificazione può anche essere riformulato in termini di un insieme di **funzioni discriminanti** $y_1(\mathbf{x}), \dots, y_c(\mathbf{x})$, tali che \mathbf{x} è associato a C_k sse

$$y_k(\mathbf{x}) > y_j(\mathbf{x}) \quad \forall j \neq k$$

Notiamo che il problema del classificatore di pattern ottimo è definito a meno di trasformazioni monotoniche nelle *funzioni discriminanti*. Sicché, p.es., la funzione discriminante che abbiamo usato, $p(\mathbf{x}|C_k)P(C_k)$ può anche essere sostituita da $\ln [p(\mathbf{x}|C_k)P(C_k)]$. Talvolta, nel caso $c = 2$, invece di $y_1(\mathbf{x})$ e $y_2(\mathbf{x})$ si preferisce usare un'unica funzione $y_1(\mathbf{x}) - y_2(\mathbf{x})$.

- Finora abbiamo assunto che minimizzare $P(\text{error})$ fosse il criterio desiderabile. Talvolta però non è così (p.es. in un caso di diagnosi medica basata su immagini è molto più grave riconoscere la presenza di un tumore dove non c'è piuttosto che l'errore opposto). La risposta a questa esigenza è definire una **matrice di perdita (loss matrix)**, i cui elementi L_{kj} rappresentano la gravità dell'errore di associare alla classe C_j un pattern che in realtà appartiene alla classe C_k . Il valore atteso delle perdite totali associate alla classe C_k sarà quindi

$$R_k = \sum_{j=1..c} L_{kj} \int_{\mathcal{R}_j} p(\mathbf{x}|C_k) d\mathbf{x}$$

mentre le perdite totali, o **rischio (risk)**, sono

$$R = \sum_{k=1..c} R_k P(C_k) = \sum_{j=1..c} \int_{\mathcal{R}_j} \{ \sum_{k=1..c} L_{kj} p(\mathbf{x}|C_k) P(C_k) \} d\mathbf{x}$$

- R è minimizzato scegliendo $\{\mathcal{R}_k\}$ in modo tale che $\mathbf{x} \in \mathcal{R}_j$ quando

$$\sum_{k=1..c} L_{kj} p(\mathbf{x}|C_k) P(C_k) < \sum_{k=1..c} L_{ki} p(\mathbf{x}|C_k) P(C_k), \quad \forall i \neq j$$

Questa formula si riduce a quella usuale se $L_{kj} = 1 - \delta_{kj}$

CAP 2. STIMA DELLA DENSITA' DI PROBABILITA'

- Il problema di stimare una funzione di densità di probabilità $p(\mathbf{x})$, dato un numero finito di punti \mathbf{x}^n , $n=1, \dots, N$, può essere risolto seguendo tre metodi:
 - a) metodo parametrico, in cui si assume una specifica forma funzionale per $p(\mathbf{x})$
 - b) metodo non-parametrico, la cui complessità cresce con N

- c) metodo semi-parametrico, il compromesso migliore
- Le NN feed-forward possono essere viste come modelli semi-parametrici per la stima di densità condizionali
 - **Metodi parametrici.** Il modello parametrico più semplice e usato è la distribuzione **normale** o **Gaussiana**.

$$p(x) = [1/(2\pi\sigma^2)^{1/2}] \exp[-(x-\mu)^2/2\sigma^2]$$

In d dimensioni abbiamo

$$p(\mathbf{x}) = \{1/[(2\pi)^{d/2}|\Sigma|^{1/2}]\} \exp[-(\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})/2]$$

dove Σ è la **matrice di covarianza** ($d \times d$). Vale

$$\boldsymbol{\mu} = \mathcal{E}[\mathbf{x}]$$

$$\Sigma = \mathcal{E}[(\mathbf{x}-\boldsymbol{\mu})(\mathbf{x}-\boldsymbol{\mu})^T]$$

Dato che Σ è simmetrica, ha $d(d+1)/2$ elementi indipendenti; $\boldsymbol{\mu}$ ha altri d elementi, sicché abbiamo in totale $d(d+3)/2$ parametri. Questi parametri sono specificati una volta noti $d(d+3)/2$ punti della distribuzione.

La quantità $\Delta^2 = (\mathbf{x}-\boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})$ è nota come **distanza di Mahalanobis** di \mathbf{x} da $\boldsymbol{\mu}$.

Le superfici $p(\mathbf{x}) = \text{cost.}$ sono delle iperellissoidi in cui $\Delta^2 = \text{cost.}$ Gli assi principali di questi iperellissoidi sono dati dagli autovalori \mathbf{u}_i di Σ , che soddisfano

$$\Sigma \mathbf{u}_i = \lambda_i \mathbf{u}_i$$

dove gli autovalori λ_i danno le varianze sulle rispettive direzioni \mathbf{u}_i .

Nel caso particolare di Σ diagonale, il numero dei parametri si riduce a $2d$. In questo caso gli iperellissoidi hanno assi allineati con gli assi coordinati, ed è possibile scomporre $p(\mathbf{x})$ nel prodotto di d Gaussiane unidimensionali

$$p(\mathbf{x}) = \prod_{i=1, \dots, d} p(x_i)$$

Un'ulteriore semplificazione si ha scegliendo $\sigma_i = \sigma$, $\forall i$, nel qual caso i luoghi a $p(\mathbf{x}) = \text{cost.}$ sono delle ipersfere.

I metodi principali per determinare i parametri ottimi del modello che fittano i dati sono la **massima verisimiglianza** (*maximum likelihood*) e l'**inferenza Bayesiana** (*Bayesian inference*).

