Tullio Tinti

Genova, 1996

Introduzione

Il connessionismo è uno dei settori in cui si articola la scienza cognitiva contemporanea. Prima di collocarlo nella giusta posizione rispetto agli altri rami di questa disciplina, è necessario chiarire cosa intendiamo con scienza cognitiva.

Il filosofo Andy Clark, aprendo il suo libro *Microcognizione*, ci offre una pittoresca definizione che adotteremo per la sua efficace sinteticità:

> La scienza cognitiva è la febbre dell'oro della mente. Tutti sono alla ricerca della mente. E ciò che è peggio tutti (ricercatori dell'IA simbolica, dell'IA subsimbolica, neuroscienziati, filosofi naturalisti e così via) sostengono che la stanno per trovare. O almeno di sapere dove cercarla [Clark 1989, 10].

Clark prosegue mettendo in luce il fatto che, contrariamente all'opinione di molti ricercatori, neppure *dove* cercare la mente è un problema risolto (tanto meno è prossima la sua "scoperta"). Questo che cosa significa? Perché non si sa *dove* cercare?

Il riferimento è ovviamente al cervello e al suo *ruolo* nella ricerca della mente. Impostata in questo modo la questione ha origini antiche, nel rapporto tra mente e corpo. Le risposte che nel tempo sono state date a questo problema costituiscono altrettante direzioni verso cui ci si può muovere alla ricerca della mente.

Una prima soluzione al "problema mente-corpo" è quella del dualismo, a cui si contrappone il monismo materialistico; secondo dualisti e materialisti, la mente va cercata rispettivamente *fuori* e *dentro* il cervello. Benché queste due strade appaiano a prima vista esaurire tutte le possibilità, in un certo senso (che ora preciseremo) una parte dei ricercatori della scienza cognitiva rifiuta entrambe le risposte. Questo fa sì che molti possano essere i punti di vista con cui avvicinarsi ai "cercatori d'oro".

Clark -ad esempio- sceglie, come (prima) discriminante tra le varie correnti della scienza cognitiva, il diverso rapporto dei ricercatori nei confronti della *folk psychology*, cioè indica come possibili strade iniziali quella di chi ritiene necessario, per spiegare l'*elaborazione mentale*, l'apparato della psicologia del senso comune e quella di chi, al contrario, ritiene fuorviante ogni diretta corrispondenza tra attività cognitiva e *folk psychology*.

Tuttavia qui preferiamo ricorrere ad un punto di vista più classico, al quale peraltro converge presto Clark, mediante cui ci si può fare strada tra le correnti della scienza cognitiva partendo proprio dal "problema mente-corpo". Esistono attualmente due impostazioni estreme, in

mezzo alle quali procede il connessionismo, e ciascuna delle due è riconoscibile in base al ruolo che il cervello si vede assegnato.

Una prima impostazione è quella del *funzionalismo computazionale*, secondo il quale il ruolo del cervello nella ricerca della mente è *minimo*. La posizione opposta è di quanti ritengono invece che il suo ruolo sia *massimo*, il cervello essendo caratterizzato da *poteri causali specifici* o da altre peculiarità non riproducibili artificialmente. E' evidente che secondo i funzionalisti, non essendo il cervello *necessario* alla mente, quest'ultima è in linea di principio ricreabile su una macchina, il cui *hardware* starebbe al *software* come il cervello umano sta alla mente ("metafora del calcolatore"). Quest'ipotesi, nota come ipotesi "forte" dell'Intelligenza Artificiale, è proprio quella contro cui si scagliano gli avversari del funzionalismo (ad esempio il filosofo John Searle [1980] ed il fisico Roger Penrose [1989]).[1] Prima di chiarire la posizione del connessionismo all'interno di questo dibattito, è importante osservare che né i funzionalisti né i critici dell'IA forte accettano senza riserve l'una o l'altra risposta di dualisti e materialisti al "problema mente-corpo"; ad esempio, presentando il funzionalismo, il filosofo Jerry Fodor scrive: «*Negli ultimi quindici anni [...] è emersa una filosofia della mente, che non è né dualistica né materialistica, e che prende il nome di funzionalismo*» [Fodor 1981, 19]. Analogamente Searle, nell'Introduzione ad un suo recente libro, scrive: «*uno degli scopi di questo lavoro è [...] criticare e superare le tradizioni filosofiche - siano esse "materialiste" o "dualiste"*» [Searle 1992, 9].

Qual è il ruolo del cervello secondo i connessionisti? Essi rispondono che è quello di *ispirare modelli computazionali* [Rumelhart 1986, 180]. E' questo concetto di "ispirazione neurale" ad essere carico di conseguenze filosofiche, come si cercherà di mostrare nel presente lavoro.

Chiediamoci prima di tutto dove finiscono i poteri causali specifici di Searle o la "metafora del calcolatore" dei funzionalisti computazionali. Secondo i connessionisti, il cervello non possiede alcun potere causale irriproducibile, *a priori*, su una macchina. Anzi, in un senso non banale il cervello è esso stesso un calcolatore (questa posizione fa sì che il connessionismo venga considerato parte della ricerca in Intelligenza Artificiale). Il fatto è che esso non lavora come i calcolatori seriali "di von Neumann" (IA *convenzionale*), costituiti da un *hardware* su cui vengono implementati i programmi. Al contrario, il cervello è esso stesso modello di un tipo diverso di calcolatori, con moltissimi processori (ciascuno più lento dei moderni microprocessori digitali) operanti *in parallelo* le cui connessioni possono modificarsi nel tempo. Gli psicologi David

[1]Naturalmente tra questi opposti vi sono molte posizioni intermedie (oltre al connessionismo) e le stesse impostazioni "estreme" si presentano articolate in numerosissime correnti di pensiero diverse. Il punto di vista qui adottato è stato scelto tenendo presente quali autori sono stati presentati nell'ambito di questo ciclo di seminari.

Rumelhart e James McClelland, intorno ai quali si è formato il cosiddetto "gruppo PDP", così commentano il concetto di "ispirazione neurale":

Il nostro desiderio è quello di sostituire alla «metafora del calcolatore» la «metafora del cervello» come modello di mente [Rumelhart 1986, 114].

In base a questa concezione, non esistono cose del tipo dello «hardware». E neppure del tipo del «software». Ci sono solo connessioni. E tutte le connessioni sono in un certo senso hardware (in quanto sono entità fisiche) e tutte sono in un certo senso software (in quanto possono essere cambiate) [Rumelhart 1986, 194].

Questo cambiamento di prospettiva costituisce una vera e propria rivoluzione nell'ambito della scienza cognitiva, perché rappresenta il passaggio da un modello del mentale basato a-criticamente sulla «*visione che la mente stessa ha della mente*» [Clark 1989, 13], cioè il punto di vista che Clark chiama "occhio della mente", ad una nuova classe di modelli la cui architettura è simile (ispirata) alla struttura cerebrale, alternativa che Clark definisce «*il punto di vista dell'occhio del cervello*» [*ibid*.].

Prima di passare all'esame dei modelli dell'"IA *connessionista*", chiediamoci come si vede la vecchia questione dualismo/materialismo "con l'occhio del cervello". I connessionisti rifiutano decisamente ogni forma di dualismo e, almeno come impostazione programmatica, si dichiarano decisamente materialisti. Tuttavia, le proprietà cognitive dei modelli connessionisti sono considerate *proprietà emergenti*, il cui status ontologico è tutt'altro che ovvio. Basti pensare che al concetto di *emergenza*, parlando di mente, ricorrono tanto un dualista interazionista come il grande filosofo Karl Popper [1977] quanto un materialista eliminativo come il filosofo Paul Churchland [1989].

1. *L'elaborazione distribuita in parallelo*

I modelli connessionisti sono, almeno in teoria, un'ampia classe di modelli computazionali, unicamente accumunati dall'"ispirazione neurale". Di fatto, tuttavia, quelli di cui ci occupiamo qui coincidono con i modelli ad elaborazione distribuita in parallelo, cioè i "modelli PDP"[2]. Questi modelli sono universalmente noti come "reti neurali artificiali" o, più semplicemente, "reti neurali".

Non è superfluo sottolineare che con "reti neurali" possiamo però intendere, anche per quanto detto finora, almeno quattro cose diverse a seconda dei contesti:

- se si parla di neurofisiologia, ci riferiamo all'architettura effettivamente esistente nel sistema nervoso degli animali (reti neurali naturali o biologiche);
- se il contesto è quello della bio-matematica, ci si riferisce a *modelli neurali*, cioè a modelli teorici che descrivono nel modo più dettagliato possibile le reti neurali naturali;
- se il contesto è quello della scienza cognitiva, e più in particolare del connessionismo (è il nostro caso), allora intendiamo *modelli neuralmente ispirati*, che sacrificano molti o moltissimi particolari biologici a favore di proprietà computazionalmente interessanti;
- se infine si parla di matematica teorica o informatica, attualmente vengono chiamati reti neurali alcuni modelli matematici nati come modelli cognitivi ma "esportati" in campi del tutto diversi, per sfruttarne le caratteristiche di elaborazione distribuita in parallelo.

Purtroppo, anche all'interno del connessionismo vi sono diversi modi di intendere l'espressione "modelli PDP". Clark ne elenca tre [Clark 1989, 162]:

a) i modelli usati effettivamente dal cosiddetto "gruppo PDP", cioè da Rumelhart, McClelland e altri;
b) tutti i modelli aventi la stessa *architettura* dei modelli usati dal "gruppo PDP";
c) tutti i modelli che, indipendentemente dall'architettura, abbiano prestazioni *qualitativamente analoghe* a quelle dei modelli (a) e (b).

Concordiamo con Clark che l'unica lettura importante è la (b), pertanto in questo lavoro useremo come sinonimi: "reti neurali" e "modelli PDP" - nell'accezione (b).

Bene, *che cosa sono le reti neurali?*

Sono insiemi di unità di *elaborazione*, ciascuna connessa ad altre, operanti *in parallelo* e caratterizzate dal fatto che ciascuna informazione da esse elaborata è *distribuita* tra le unità. Da questa prima definizione è già possibile vedere perché si parli di modelli ad *elaborazione distribuita in parallelo*. Prima di esaminare maggiormente questa definizione, è necessario chiarire

se stiamo parlando di unità *fisiche* (cioè effettivamente, materialmente, esistenti) o *virtuali* (cioè esistenti solo a qualche livello di descrizione).

La risposta è che le reti neurali possono essere reali dispositivi fisici oppure normali programmi per normali computer (seriali) *descrivibili* come reti (ad elaborazione parallela). Nel primo caso si parla di *realizzazione fisica* del modello, nel secondo di *reti simulate*.

La realizzazione fisica può avvenire mediante circuiti elettronici, con amplificatori operazionali come unità di elaborazione e cavi, resistori e condensatori come connessioni [Tank 1988]; o addirittura mediante sistemi ottici, con superfici semi-trasparenti ciascun punto delle quali rappresenta la connessione tra due unità [Abu-Mostafa 1987]. La realizzazione fisica delle reti neurali è tuttavia ancora agli albori della ricerca.

La maggior parte dei modelli PDP oggi esistenti sono *simulati* su calcolatori tradizionali (e *non* su computer paralleli → vedi Scheda 1). In questo caso l'unica cosa che esiste è il programma. I programmi di simulazione delle reti sono programmi come gli altri, scritti in linguaggi di programmazione solitamente molto semplici. Tuttavia, a differenza degli altri programmi usati dai ricercatori dell'IA, *il modo ottimale per descriverli è farlo come se fossero reti neurali fisicamente realizzate*. Questo fatto non verrà ulteriormente ribadito e si darà per scontato che unità di elaborazione e connessioni vengano pensate come semplici variabili all'interno di un programma (non parallelo) per calcolatore (non parallelo), cioè che per rete si intende una *rete simulata*.

Scheda 1. *Macchine e programmi paralleli*

Si fa generalmente molta confusione quando si parla di parallelismo e di calcolatori operanti in parallelo. Poiché una delle caratteristiche più interessanti delle reti neurali è l'elaborazione in parallelo, è qui necessario chiarire almeno a grandi linee che cosa si intende con macchine e programmi operanti in parallelo.

Per *calcolo parallelo* si intende ogni tipo di elaborazione in grado di affrontare contemporaneamente le singole parti in cui può essere diviso un problema complesso. Poiché l'elaborazione elettronica è effettuata da unità fisiche di elaborazione (i processori) opportunamente programmate (mediante il *software*), ci sono almeno due strade per affrontare il calcolo parallelo: costruendo calcolatori con *hardware* parallelo oppure realizzando programmi paralleli (che, almeno in teoria, possono girare sia su macchine seriali che parallele). Nessuna di queste due strade

[2] "PDP" sta per *Parallel Distributed Processing.*

riguarda *direttamente* l'elaborazione in parallelo delle reti neurali, le quali vengono di norma *simulate* su calcolatori seriali, mediante programmi seriali.

Per quanto riguarda il *software* parallelo, praticamente è ancora tutto da fare. Solo molto recentemente hanno cominciato a diffondersi programmi per il calcolo parallelo, ma i progressi più significativi degli ultimi decenni sono venuti dalla costruzione di sistemi con *hardware* parallelo.

Nel campo dell'*hardware*, esiste un'intera gamma di architetture parallele differenti. Qui basterà ricordare che ad un estremo di questa gamma si colloca l'architettura del calcolatore sequenziale comune (di von Neumann), caratterizzato dal numero minimo di processori (uno!) con una grande potenza di calcolo; all'altro estremo vi sono i calcolatori ad alto parallelismo, in cui un numero elevatissimo di processori di scarsa potenza compiono contemporaneamente la stessa istruzione su dati differenti (architettura "SIMD": *Stessa Istruzione / Molteplici Dati*). Il più famoso calcolatore di questo tipo è la Connection Machine di Daniel Hillis, dotata di ben 65536 processori [Hillis 1987].

A metà strada si trovano calcolatori con relativamente poche unità di elaborazione, ciascuna con una potenza di calcolo piuttosto grande. Di questo tipo sono gli «elaboratori vettoriali», i cui processori hanno accesso a dati contenuti in una memoria comune, ma eseguono contemporaneamente istruzioni diverse (architettura "MIMD": *Molteplici Istruzioni / Molteplici Dati*). Ognuna di queste architetture ha vantaggi e svantaggi, in termini di velocità/potenza e costo/prestazioni [Corcoran 1991].

Torniamo ora all'elaborazione distribuita in parallelo. L'attività "computazionale" del cervello è, secondo i connessionisti, l'esempio più evidente di questo tipo di elaborazione.

I neuroni del cervello sono cellule molto ramificate che ricevono segnali elettro-chimici le une (*efferenti*) dalle altre (*afferenti*). Questi segnali hanno per lo più la funzione di stimolare o inibire la scarica delle cellule efferenti. La natura (elettrica o chimica) e la frequenza dei segnali, da sole, costituiscono l'intero spettro di differenziazione delle informazioni elaborate dal cervello.

Questa apparente "semplicità" del sistema nervoso è alla base dell'idea che il cervello sia un calcolatore biologico,

molto più semplice di quanto la nostra vanità avesse sperato o il nostro intelletto temuto [Llinas 1984, 198].

In realtà non dobbiamo farci trarre in inganno. La struttura che abbiamo appena descritto è il sistema complesso *funzionalmente più complesso* che attualmente si conosca. Tuttavia, la relativa "semplicità" architettonica dei neuroni e delle loro connessioni, le sinapsi, è proprio ciò che consente la metafora connessionista: i neuroni possono essere pensati come unità di elaborazione e le sinapsi come le connessioni tra queste unità. Da questo punto di vista, allora, non c'è dubbio che si tratti di elaborazione *in parallelo*: molti processori lavorano contemporaneamente e in maniera fittamente interconnessa; inoltre, poiché nessuna informazione è disponibile tutta in una volta per qualche neurone,[3] allora le informazioni vere e proprie sono *distribuite* tra le unità di elaborazione.

Nelle reti neurali avviene lo stesso tipo di elaborazione. Ogni unità è caratterizzata da un valore numerico chiamato *stato di attivazione* o *scarica*. Nel programma di simulazione della rete questi valori compaiono sotto forma di una variabile in funzione del tempo (o meglio, di una variabile *t* che *simula* il tempo). Le connessioni tra le unità non vengono rappresentate direttamente, ma nel programma compare un'altra variabile, una per ogni connessione, chiamata *peso sulla connessione*.[4] I pesi non sono funzioni del tempo e si modificano solo durante la procedura di *apprendimento* (che fa parte del programma di simulazione). Come viene elaborata l'informazione? Così *come* nel cervello e *diversamente* dai calcolatori digitali, nelle reti neurali non ci sono memorie di dati a cui le unità di elaborazione accedono per "manipolare le informazioni": l'unica "informazione" che si propaga è la scarica, opportunamente pesata nel passaggio dalle unità afferenti a quelle efferenti, e l'unica "elaborazione" che avviene è l'aggiornamento, nel tempo, dello stato di attivazione delle unità.

Siamo ora in grado di esaminare l'effettiva struttura delle reti. Le componenti di una rete sono [Rumelhart 1986, 81-91]:

1. n unità di elaborazione, rappresentate graficamente (nei diagrammi usati per *descrivere* la rete) mediante cerchi;
2. lo stato di attivazione (o scarica) delle unità; in forma vettoriale: $\mathbf{s}(t) = (s_1(t), ..., s_n(t))$;
3. le connessioni, rappresentate graficamente mediante frecce, e i pesi sulle connessioni, indicati con w_{ij} , che formano una matrice **W**;
4. l'ingresso-netto nelle unità; in forma vettoriale: $\mathbf{net}(t) = \mathbf{W}\ \mathbf{s}(t)$;

[3]Nella letteratura connessionista chi ritiene che le cose stiano diversamente viene spesso accusato di credere nel "neurone della nonna", cioè in cellule in grado, da sole, di manipolare informazioni molto complesse come quelle necessarie e sufficienti all'identificazione di una persona (ad esempio la nonna).

[4]In matematica si intende per *peso* un coefficiente, cioè un fattore moltiplicativo. *Pesare* un valore significa cioè moltiplicare tale valore per qualche costante (chiamata appunto peso). Per esempio, la somma pesata di A più B è la seguente: $\alpha A + \beta B$, dove α è il peso di A e β è il peso di B.

5. una regola di aggiornamento;
6. una regola di apprendimento.

1. *Unità di elaborazione*

Sono la modellizzazione dei *neuroni*, e possono essere pensate come rappresentanti: tratti, lettere, parole, concetti o semplicemente elementi astratti. Si dividono in:

- unità (visibili) di entrata;
- unità (visibili) di uscita;
- unità nascoste.

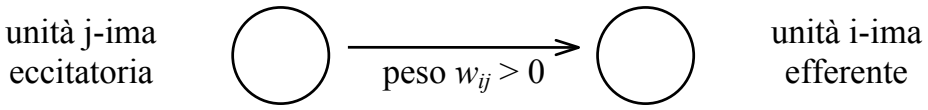
2. *Stato di attivazione o scarica*

Nei modelli PDP lo stato di attivazione e la scarica dei neuroni vengono rappresentati (nella maggioranza dei casi) congiuntamente. In ciascun istante t ogni unità è così caratterizzata da un certo valore che costituisce sia l'attivazione di quell'unità che la scarica di quell'unità verso le unità a cui afferisce. Nell'istante successivo $t + 1$, poiché ogni unità, oltre che scaricare, riceve segnali dalle unità di cui è efferente, il suo stato viene aggiornato in funzione della scarica ricevuta ("ingresso-netto").

3. *Pesi sulle connessioni*

Sono la modellizzazione delle *forze sinaptiche*. Se nel modello si hanno connessioni solo eccitatorie o inibitorie, allora il peso assume valori positivi (connessione eccitatoria), negativi (connessione inibitoria) o nulli (nessuna connessione diretta), ma se ci sono molti tipi di connessioni, allora si avrà una matrice di pesi *per ogni tipo di connessioni*.

Nel caso più semplice (connessioni eccitatorie/inibitorie), scriviamo per esempio:



4. *Ingresso-netto*

Ogni unità che afferisce ad un'unità contribuisce con un ingresso che è dato dal prodotto del peso sulla connessione per la scarica dell'unità afferente:

all'istante t l'unità j-ima riceve dall'unità i-ima un ingresso = $w_{ji}\, s_i(t)$.

L'ingresso-netto è la somma degli ingressi. Si ha quindi che $net_j(t) = \sum_i w_{ji}\, s_i(t)$, cioè:

$$\mathbf{net}(t) = \mathbf{W}\ \mathbf{s}(t)\ .$$

5. *Regola di aggiornamento*

Occorre una regola che stabilisca, per ogni unità j-ima che al tempo t si trova nello stato di attivazione $s_j(t)$ e riceve un ingresso-netto $net_j(t)$, quale sarà lo stato di attivazione, e quindi la scarica, dell'unità nell'istante successivo $t + 1$. Questa regola è di solito una funzione dell'ingresso-netto, anche se possono esserci casi più complicati. Un classico esempio è la funzione binaria a soglia: l'unità efferente scarica il valore 1 se il suo ingresso-netto è sopra una certa soglia, non scarica (stato di attivazione = 0) se l'ingresso-netto è sotto la soglia.

6. *Regola di apprendimento*

L'apprendimento consiste in una regola di modificazione di **W**. La regola di apprendimento più famosa è quella dovuta allo psicologo Donald Hebb, basata sull'idea che l'attivazione contemporanea di due unità aumenta la forza della loro connessione (e viceversa). Purtroppo non solo questa è l'*unica* forma di apprendimento "neuralmente ispirata" dei modelli PDP, ma l'individuazione del suo corrispettivo biologico (le «sinapsi di Hebb») è ancora molto controversa. La regola di apprendimento più usata dal "gruppo PDP" è la cosiddetta «regola delta generalizzata», basata sulla *minimizzazione automatica dell'errore* in seguito ad esempi del tipo: <ingresso, uscita corretta>. Anche se nel sistema nervoso nessun processo di minimizzazione dell'errore è stato individuato dalle neuroscienze, esso è perlomeno compatibile con la nostra concezione dell'evoluzione naturale [Churchland 1992, 196-204].

Prima di passare all'esame qualitativo delle reti neurali, desideriamo rivolgerci ancora una volta al programma che le simula, per sottolineare di nuovo in che cosa consiste effettivamente l'elaborazione distribuita in parallelo.

Di fatto, quello che fa un tipico programma di simulazione di rete neurale è:

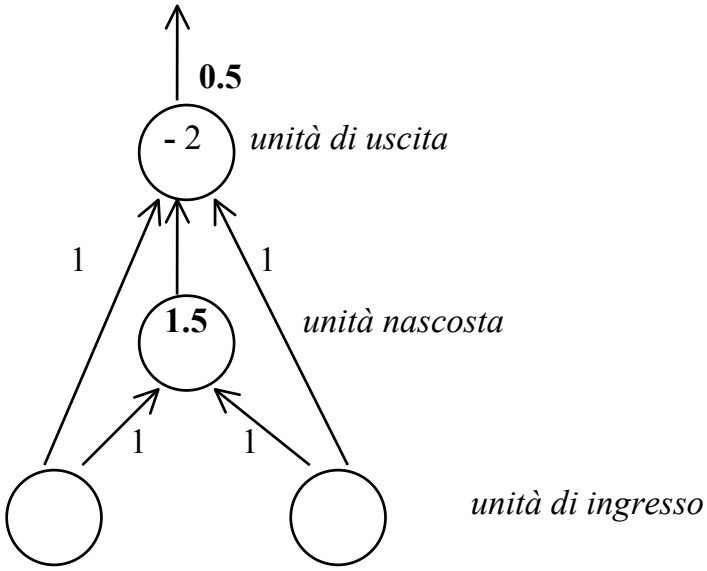
1. richiedere all'utente i valori di ingresso da elaborare;
2. assegnare allo stato delle unità di entrata i valori di ingresso e allo stato delle altre unità e a **W** valori a caso;

3. calcolare il valore dell'ingresso-netto ricevuto da ogni unità efferente, cioè la somma pesata delle scariche delle unità afferenti;
4. aggiornare lo stato di ogni unità secondo la funzione di aggiornamento in base all'ingresso-netto nell'unità;
5. visualizzare il valore della scarica delle unità di uscita (ad "uso e consumo" dell'utente).
6. Se nel programma esiste una procedura di *apprendimento*, viene richiesta all'utente una valutazione della differenza tra il risultato appena visualizzato e il risultato desiderato (questa differenza viene chiamata *delta*); in base a tale valutazione il programma modifica i valori di **W**;
7. infine tutta la procedura si ripete.

Scheda 2. *Esempio: il problema della disgiunzione esclusiva*

Consideriamo un esempio molto famoso di rete neurale, la rete in grado di risolvere il problema della disgiunzione esclusiva (*aut*). Com'è noto, questa è la funzione logica più simile all'irrinunciabile concetto ordinario di «diversità»: vale 1 se gli ingressi sono diversi, cioè (1,0) oppure (0,1), e 0 se gli ingressi sono uguali, cioè (0,0) oppure (1,1). Si tratta di un esempio famoso perché le prime reti, prive delle cosiddette unità nascoste, non riuscivano a calcolare questa funzione. E' infatti indispensabile (per ragioni matematiche) la presenza di *almeno un'unità* collocata tra le unità di entrata e le unità di uscita. Il testo del "gruppo PDP" del 1986 è, anche a questo proposito, un'opera fondamentale, perché in esso veniva esposta per la prima volta al grande pubblico una regola di apprendimento, nota come «regola delta generalizzata», che consente alle reti (con almeno un'unità nascosta) di aggiustare automaticamente i pesi in base all'errore commesso (*delta*).

Vediamo solo un esempio numerico, dei tanti possibili, di rete che risolve il problema. La rete sia composta da quattro unità: due di entrata, una nascosta e una di uscita. La regola di attivazione sia una funzione binaria a soglia (stato di attivazione = 0 se l'ingresso-netto è sotto la soglia, 1 se sopra la soglia) e la soglia di ciascuna unità sia indicata nel cerchio rappresentante l'unità medesima. L'ingresso-netto è, come si ricorderà, la somma degli ingressi (ingresso = scarica afferente × peso sulla connessione). Il compito dell'unità nascosta è inibire la scarica dell'unità di uscita quando entrambe le unità di ingresso sono attivate [Rumelhart 1986, fig.5.2 p.204]:



Come si può facilmente vedere, con questi valori di soglia e questa distribuzione di pesi, la rete riesce a calcolare correttamente la funzione della disgiunzione esclusiva.

Per esempio, se l'ingresso è (0,1): l'unità nascosta riceve un unico ingresso (dall'unità di entrata di destra), corrispondente a 1 (perché il peso è 1). Questo valore è anche il suo ingresso-netto, che risulta inferiore alla soglia (1.5). In questo caso l'unità nascosta non scarica. All'unità di uscita arriva come ingresso-netto la scarica dell'unità di entrata di destra (pesata da 1), superiore alla sua soglia. Lo stato di attivazione dell'unità di uscita, cioè il risultato finale dell'elaborazione, è allora 1, come doveva essere.

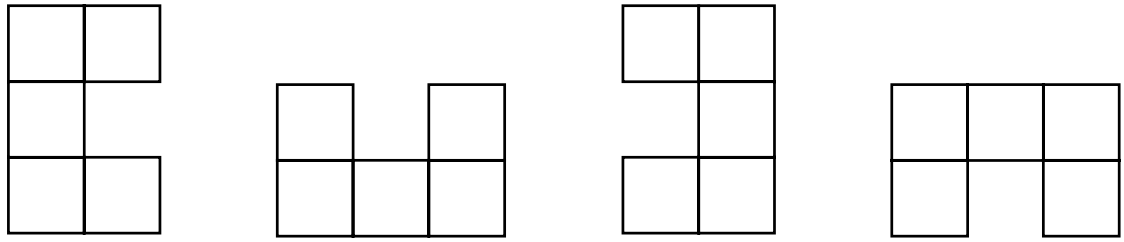
Se l'ingresso da elaborare è invece (1,1): entrambe le unità di entrata scaricano e l'ingresso-netto nell'unità nascosta diventa = (1×1) + (1×1) = 2, superiore alla soglia. In questo caso l'unità di uscita riceve gli ingressi da tutte le altre unità e il suo ingresso-netto risulta = (1×1) + (1×1) + (-2 × 1) = 0, inferiore alla soglia (0.5). Stavolta lo stato di attivazione dell'unità di uscita, cioè il risultato finale dell'elaborazione, è 0, come doveva essere.

C'è una grande differenza tra quello che i sostenitori dei modelli PDP promettono e quello che effettivamente le loro reti riescono a fare. Come dice Clark, il programma di ricerca connessionista

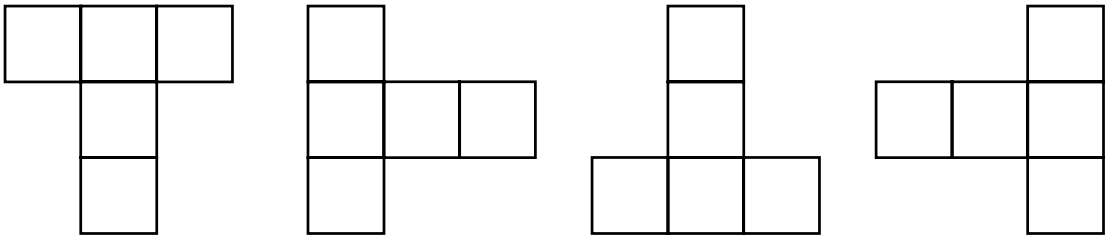
si propone di offrire niente di *meno* che un nuovo modello computazionale della mente. Purtroppo non offre nulla di *più* che alcune intuizioni sulla natura e la potenza di questi modelli [Clark 1989, 7].

E' interessante fare un esempio concreto di cosa fanno le reti neurali. Consideriamo il cosiddetto "problema T-C", cioè il semplice problema di distinguere una lettera (molto stilizzata!) dall'altra [Rumelhart 1986, 236-41].

La rete usata per risolvere questo problema è costituita da un insieme molto ampio di recettori, collegati alle unità nascoste. Ciascuna unità nascosta riceve segnali da una zona quadrata di 3 × 3 recettori, chiamata *campo recettivo*. Tutti i campi recettivi della rete sono uguali tra loro e sono parzialmente sovrapposti in modo sistematico. Scopo della rete è determinare se sull'insieme di recettori si trova una T oppure una C, nei quattro orientamenti consentiti. Entrambe queste lettere sono composte da cinque quadratini, ciascuno in grado di attivare un recettore; la C si può presentare nei seguenti modi:



e la T può essere presente in una delle quattro forme:



Scopo dell'apprendimento è stato trovare i valori da assegnare a ciascun recettore del campo recettivo, in modo tale da discriminare la presenza di una lettera piuttosto che l'altra.

Mediante il metodo della regola delta generalizzata sono state trovate varie soluzioni a questo problema. Vediamo la cosiddetta soluzione "centro *on* - periferia *off*". Il campo recettivo è il seguente:[5]

-1	-1	-1
-1	2	-1
-1	-1	-1

[5]I valori riportati (-1 e 2) sono approssimati.

E' evidente che se la lettera cade nel campo recettivo di qualche unità con un solo quadratino, realizzando -1, l'unità non sarà in grado di distinguere la lettera. Neppure con due quadratini allineati ci sarà discriminazione, se i due quadratini si trovano lungo il perimetro del campo recettivo, perché il valore di attivazione sarà -2 sia per la C che per la T. D'altra parte, se la lettera centra in pieno il campo recettivo, attivando il recettore centrale, il valore sarà di nuovo -2 (quattro recettori del perimetro più il recettore centrale):

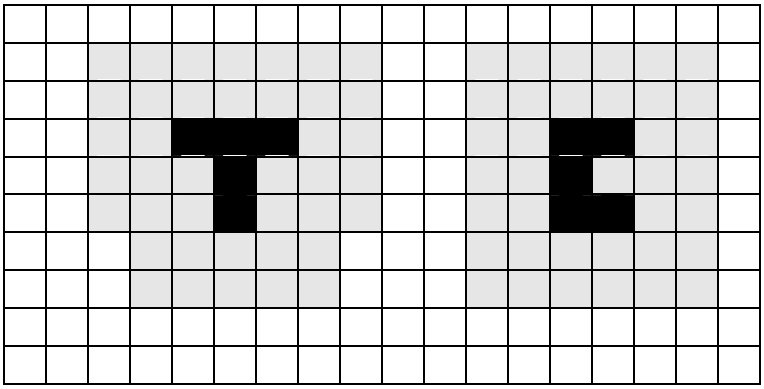
-1	-1	-1
-1	2	-1
-1	-1	-1

-1	-1	-1
-1	2	-1
-1	-1	-1

-1	-1	-1
-1	2	-1
-1	-1	-1

Tuttavia, se la lettera copre un unico recettore periferico più il recettore centrale del campo recettivo, realizzando il valore +1, allora si tratterà certamente di una T, perché in nessun orientamento la C può ottenere questa attivazione. Poiché i campi sono tutti parzialmente sovrapposti, ci sarà sempre almeno un campo in cui la T copre un recettore del perimetro e il recettore centrale, per cui *se un campo recettivo attiva la propria unità nascosta con il valore* +1, *allora la lettera è sicuramente una T.* Poiché le unità nascoste hanno tutte una soglia tale da scaricare solo se attivate da un segnale ≥ 1, allora la rete scaricherà solo in presenza di una T.

Un'altra soluzione al problema T-C appresa mediante la regola delta generalizzata è la seguente. I valori di soglia ed i pesi si sono fissati in modo tale che tutte le unità nascoste sono attive, a meno di venir disattivate. Se almeno un recettore del campo di un'unità viene attivato (cioè coperto dal quadratino della lettera), *tutta l'unità si spegne.* Questa soluzione discrimina le lettere, perché i cinque recettori attivati dalle C sono collegati complessivamente a 20 unità nascoste, mentre i recettori coperti dai cinque quadratini delle T sono collegati complessivamente a 21 unità nascoste (si ricordi che i campi recettivi sono tutti parzialmente sovrapposti in maniera sistematica):



Questa particolare soluzione trasforma la rete in un *rilevatore di compattezza*. Funziona perché la C è più compatta della T.

In tutti i casi le soluzioni al problema T-C sono state raggiunte presentando *migliaia* di volte al sistema le otto configurazioni ammesse.

Già questo esempio ci fornisce un'idea di quanto sia lontano il programma connessionista dal "trovare la mente". Benché le soluzioni trovate automaticamente dalla rete risultino compatibili con le informazioni provenienti dalla biologia (ad esempio le *cellule gangliari retiniche* sono contraddistinte da un'organizzazione del tipo "centro *on* - periferia *off*" [Churchland 1992, 84-5]), resta il fatto che questo modello non distingue *tutte* le lettere T da *tutte* le lettere C, ma ne riconosce una versione semplificata in maniera impressionante. Anche limitandoci alle lettere maiuscole in stampatello, basta sfogliare una rivista per incontrare decine di caratteri diversi: T, **T**, *T*, T, T, T e C, *C*, **C**, C, C, C (naturalmente in questi 10 anni ci sono stati molti *progressi*, ma nessuno *rivoluzionario*).

2. *Vantaggi e limiti dell'ispirazione neurale*

In questo paragrafo si cercherà di dare un'idea schematica della potenza e dei problemi dei modelli neuralmente ispirati. E' evidente che già in linea di principio l'idea dell'ispirazione neurale è un azzardo: si tratta di fornire un modello di funzioni che (per ipotesi) sono elaborate dal cervello in modo *ampiamente distribuito*, utilizzando però un numero di unità di elaborazione circa dieci ordini di grandezza più piccolo del numero di neuroni e un numero di connessioni probabilmente circa *dodici ordini di grandezza* più piccolo del numero delle sinapsi di un cervello. Ora: per quanto si abbia poca dimestichezza con i grandi numeri è palese che si tratta di un salto di livelli pazzesco, tale da compromettere in partenza qualunque pretesa di rigore scientifico. Tuttavia, già l'idea stessa dell'ispirazione neurale (così ovvia dal punto di vista dell'"occhio del cervello"), è sufficiente a fornire i modelli PDP di una serie di interessanti proprietà, assolutamente insperabili nell'IA convenzionale.

Dovendo evidenziare la novità più importante delle reti neurali rispetto ad intere generazioni di programmi convenzionali, dovremmo senz'altro indicare la loro *flessibilità* nel trattare dati scorretti o incompleti. Probabilmente si tratta della qualità fondamentale del sistema nervoso e non stupisce che sia la prima caratteristica ereditata da modelli neuralmente ispirati.

Il problema più grave è, com'era prevedibile, la riproduzione delle qualità cognitive più lontane dalla base biologica: quelle, per intenderci, soprannominate "*di alto livello*" e implementate sui calcolatori, con discreti risultati, mediante i programmi dell'IA convenzionale.

Per esaminare la potenza dei modelli PDP considereremo il simpatico esempio del "quartiere malfamato" di McClelland [Rumelhart 1986, 59-63; Clark 1989, 121-8]. Si tratta di una rete in cui alcune unità rappresentano gli individui appartenenti a due bande giovanili, altre rappresentano i loro nomi, altre le loro professioni, altre le loro età, altre ancora il loro grado di istruzione ed infine due le bande di appartenenza. Ciascuna unità del primo gruppo (individui) ha *connessioni eccitatorie* con *un'unica unità per ognuno degli altri gruppi*, cioè con le unità che rappresentano le proprietà dell'individuo in questione. All'interno dei gruppi, invece, tutte le unità hanno *connessioni mutualmente inibitorie* (in modo che se all'interno del gruppo che rappresenta le età viene per esempio attivata l'unità «20 anni», vengano contemporaneamente inibite le unità: «30 anni», «40 anni», ecc.). Vediamo quello che succede.

Attivando come ingresso l'unità che rappresenta il nome «Rick» (supponendo che tutti abbiano un nome diverso), essa inibisce tutti gli altri nomi ed eccita l'unità che rappresenta Rick nel gruppo degli individui. Questa, a sua volta, inibisce tutte le altre unità del suo gruppo ed eccita le

unità che rappresentano le varie caratteristiche di Rick, ciascuna delle quali inibisce le "rivali" del proprio gruppo. In conclusione: la rete fornirà in uscita la "descrizione" di Rick, facendo scaricare le unità che rappresentano le sue proprietà (e solo quelle). Questa "descrizione" finale viene definita *rilassamento* della rete.

- *Memoria a indirizzabilità per contenuto*

La semplice architettura di questa rete è sufficiente a fornire la giusta descrizione degli individui anche se ne ricordiamo una descrizione parziale (e non il nome). Per esempio, se Rick è l'unico Squalo di 20 anni, attivando come ingresso l'unità «Squali» nel gruppo delle bande e quella «20 anni» nel gruppo delle età, il risultato sarà il medesimo che si ha attivando «Rick». Questo tipo di memorizzazione è detta *indirizzabilità per contenuto* ed è contrapposta a qualsiasi tipo di indirizzabilità per *etichette* precise.

- *Degrado graduale*

Poiché la scarica inibitoria che un'unità esercita sulle altre (appartenenti al suo stesso gruppo) è *proporzionale* al numero di unità che la eccitano, la rete ha l'importantissima caratteristica di rispondere sensatamente agli ingressi che contengono degli errori.

Per esempio, introducendo nella rete di McClelland una descrizione parziale (niente nome) e con un'informazione *sbagliata*, verranno attivate parecchie unità del gruppo degli individui. Ma l'unità che rappresenta l'individuo che *più si avvicina* alla descrizione riceverà un ingresso-netto maggiore delle altre e tenderà ad inibire le rivali con *maggior forza* di quanto non facciano tali unità nei suoi confronti. In questo modo essa prevarrà ed attiverà la relativa unità del gruppo dei nomi. Così, nella "descrizione" finale, sarà attiva (nel gruppo dei nomi) l'unità del nome dell'individuo che *più* si avvicina alla descrizione parzialmente sbagliata.

- *Assegnazione per difetto*

Come fa la rete ad attivare le unità giuste nel caso di descrizioni parziali? Ad esempio, se diamo come ingresso il nome (per esempio «Lance») e la banda di un individuo e vogliamo sapere qual è il suo mestiere, cosa succede? L'unità del gruppo delle bande attiverà tutte le unità del gruppo degli individui appartenenti a quella banda. A loro volta queste unità attiveranno tutte le varie proprietà dei rispettivi individui. Fortunatamente, però,

> se la maggior parte di coloro che condividono le proprietà *conosciute* di Lance hanno in comune anche qualche altra proprietà (cioè, se c'è una regolarità reale) allora l'attivazione diffusa da queste unità si combinerà per attivare per Lance l'unità che rappresenta l'ulteriore proprietà in questione [Clark 1989, 127].

Così, se la maggior parte degli appartenenti alla banda di Lance sono, per esempio, scassinatori, l'unità «scassinatore» si attiverà come assegnazione "per difetto" (cioè in mancanza di dati relativi) a Lance.

- *Generalizzazione flessibile*

Poiché in ciascun gruppo dominano le unità maggiormente pertinenti ad un ingresso parziale, è possibile avere la descrizione di un "individuo-medio" avente una certa proprietà anche se nessuno corrisponde davvero a quella descrizione. Così, se la *maggioranza* degli appartenenti agli Squali sono scassinatori, se la *maggioranza* di essi ha 20 anni e la *maggioranza* di essi ha un grado inferiore di istruzione, allora attivando come ingresso solo l'unità «Squali», si attiveranno anche «scassinatore», «20 anni» e «istruzione inferiore» - anche se *nessuno* è uno Squalo scassinatore di 20 anni con grado inferiore di istruzione.

E' questa la qualità più affascinante delle reti, spesso resa con immagini molto forti («la rete "si è fatta un'idea" dello Squalo tipico», «la rete "ha in mente il prototipo" dello Squalo», ecc.), ma che in realtà è la semplice conseguenza della particolare struttura del modello: *eccitazione tra classi di elementi e mutua inibizione tra elementi della stessa classe*.[6] In altre parole la rete ha memorizzato quello che è stata congeniata per memorizzare: una *rappresentazione distribuita*, e non *localizzata*, degli individui appartenenti alla banda degli Squali.

Veniamo ora ai limiti dei modelli PDP. Nell'esaminare alcuni problemi cercheremo di integrare quanto detto finora con alcune idee della *filosofia* del connessionismo, come per esempio quelle relative al concetto di *emergenza*.

Molte critiche che sono state fatte ai modelli connessionisti non tengono conto in nessun modo del fatto che lo studio dell'elaborazione distribuita in parallelo è appena agli inizi. Molti problemi verranno certamente risolti con l'esperienza "sul campo". Altre critiche, invece, puntano a minare alla base l'*idea* del connessionismo. Queste sono molto più interessanti di quelle di chi "spara sulla crocerossa" e sono quelle che verranno presentate, in rapida rassegna, qui di seguito. Non si tratta di un elenco completo ma solo di una carrellata sufficiente per avere un'idea dei problemi.

- *Problemi "tecnici"*

[6]E' interessante il fatto che l'analisi della distribuzione dei pesi di una rete (*analisi di cluster*) rivela spesso che la rete stessa, nell'apprendimento mediante esempi, tende a strutturarsi in questo modo, cioè raggruppando gli elementi mutualmente inibitori in classi mutualmente eccitatorie (*clustering*), anche senza alcuna istruzione specifica, a questo riguardo, nell'algoritmo di apprendimento del programma.

Un semplice esempio di problema "tecnico" molto insidioso riguarda le modalità dell'apprendimento. Non solo gli algoritmi di apprendimento attualmente utilizzati non sono neuralmente ispirati, non solo richiedono una valutazione esterna (da parte dell'utente) della prestazione della rete, non solo incontrano difficoltà matematiche (il problema dei "minimi locali"), ma -soprattutto- richiedono un numero esorbitante di esempi. Pensiamo alla rete che risolve il "problema T-C": essa apprende la differenza tra le due lettere dopo 5000 o 10000 esempi. Dopo tale incredibilmente lento addestramento, i pesi saranno fissati in modo tale da risolvere il problema. Tuttavia, se volessimo insegnare a tale rete *anche* qualcos'altro, essa dovrebbe *imparare da capo* il nuovo problema. Cioè, una volta appreso come risolvere il problema A, le reti non possono imparare a risolvere il problema B *senza dimenticare* come risolvere A. Esse devono imparare un problema *per loro* completamente nuovo, che solo agli occhi dell'utente umano è costituito da A + B. Questo porta ad una vera e propria *esplosione numerica* del numero di esempi necessari per imparare a risolvere problemi complessi, la qual cosa -com'è evidente- non si verifica nel caso dell'apprendimento degli esseri umani.

- *La critica del costruttivismo anti-rappresentazionale*

Una famosa critica "di principio" viene dagli epistemologi "costruttivisti" e riguarda il fatto che le reti neurali siano sistemi rappresentazionali. Nella filosofia del connessionismo il concetto di *rappresentazione* ha un ruolo chiave: una mente (intesa come proprietà emergente di una rete neurale) immersa ed agente nel mondo, deve farsene una *rappresentazione*. L'esempio paradigmatico scelto dal "gruppo PDP" è quello del gioco. Si tratta di una situazione in cui il sistema (la mente) deve poter rispondere a due domande precise: «che cosa devo fare?» e «che cosa accadrà una volta eseguita una data azione?». Una parte del sistema deve dunque *agire in base ai dati* e un'altra parte deve invece *simulare azioni* per fare delle previsioni:

> Questo secondo modulo del sistema potrebbe essere considerato un modello mentale degli eventi esterni. [...] Se il modello del «mondo» è ragionevolmente fedele, è possibile usarlo per scoprire le varie conseguenze delle nostre azioni, proprio come se le eseguissimo davvero [Rumelhart 1986, 295-6].

(Il concetto di rappresentazione, come vedremo, torna anche nel tentativo di rispondere al *problema del controllo*.) Questa visione di un mondo dato, nel quale il sistema mentale è immerso ed agisce e dal quale riceve informazioni e retroazioni, è criticata dal biologo costruttivista Francisco Varela, che contro *ogni* concetto di «modello del mondo» propone il suo concetto di *chiusura operazionale*:

Il concetto di chiusura operazionale è [...] un modo per specificare classi di processi che, nel loro funzionamento, si richiudono su se stessi a formare reti autonome. [...] Il punto fondamentale è che tali sistemi non funzionano attraverso rappresentazioni. Invece di *rappresentare* un mondo indipendente, essi *producono* un mondo come dominio di distinzioni inscindibile dalla struttura incarnata dal sistema cognitivo [Varela 1991, 170-1].

La critica anti-rappresentazionalista è una questione di principio sulla concezione della realtà e del sistema <soggetto-mondo>. Benché le ragioni dei costruttivisti siano molto forti, occorre ricordare che molte teorie contemporanee, elaborate da studiosi che lavorano ai confini tra linguistica, psicologia cognitiva, filosofia e IA, stanno cercando di *riformare* il concetto di rappresentazione, senza per questo volerlo abbandonare [Marconi 1995, 448-9].

- *La critica degli anti-riduzionisti*

Il "gruppo PDP" affronta esplicitamente l'accusa di fare «esercizio di riduzionismo». La risposta è molto diretta e convincente, anche se chiama in causa il problematico concetto di *emergenza*. Rumelhart e colleghi sembrano anche distinguere la *comprensione* dalla *spiegazione* dei fenomeni cognitivi. Mentre i riduzionisti cercano di *spiegare* il mentale riducendolo al cervello, i connessionisti cercano solo di *capirlo*:

Siamo consapevoli del fatto che ai diversi livelli di organizzazione emergono nuovi e utili concetti. Stiamo semplicemente cercando di *capire* l'essenza della cognizione, come proprietà emergente dalle *interazioni* di unità connesse in reti.

Noi crediamo certamente ai fenomeni emergenti, nel senso di fenomeni che non potrebbero essere mai capiti o predetti da uno studio degli elementi inferiori isolati. [...] Per esempio, potremmo non conoscere i diamanti partendo dallo studio di atomi isolati; potremmo non capire la natura dei sistemi sociali, partendo dallo studio di individui isolati; e potremmo non capire il comportamento delle reti di neuroni, partendo dallo studio di neuroni isolati. [...] Ciò comunque non indica che la natura degli elementi di livello inferiore non sia rilevante per i livelli superiori di organizzazione; al contrario, il livello superiore, a quanto crediamo, va capito in primo luogo attraverso lo studio delle interazioni tra le unità di livello inferiore. [...] *Possiamo* capire perché i diamanti sono duri, non come fatto isolato, ma perché capiamo come gli atomi di carbonio possono allinearsi a formare un reticolo perfetto. E' questa una caratteristica dell'aggregato, non dei singoli atomi, ma le caratteristiche degli atomi sono necessarie per capire il comportamento dell'aggregato [Rumelhart 1986, 177-9].

Il vero carattere della scienza cognitiva consiste nel tentativo di *spiegare* i fenomeni mentali attraverso la *comprensione* dei meccanismi che sono alla loro base [Rumelhart 1986, 168 - corsivo mio].

- *La critica di Fodor e Pylyshyn*

Alcuni anni fa, nel numero 28 della rivista «Cognition» (1988), apparvero, uno di seguito all'altro, due articoli "di fuoco" contro il connessionismo. Il primo e più famoso era firmato da

Jerry Fodor e Zenon Pylyshyn, il secondo da Alan Prince e Steven Pinker. Entrambi gli articoli sono affrontati da Clark [1989, capp.8-9]. Qui basterà ricordare che la critica di Fodor e Pylyshyn verte intorno al problema della *composizionalità*: le teorie cognitive classiche spiegano la sistematicità del pensiero postulando rappresentazioni interne con la stessa struttura sintattica delle proposizioni del linguaggio naturale. Questo comporta che tali rappresentazioni interne siano *composte* da parti che determinano il significato delle espressioni in cui compaiono. Ma ciò significa che la composizionalità richiesta è di livello concettuale, cosa *postulata* dai cognitivisti come Fodor e Pylyshyn ma *negata* dai connessionisti, secondo i quali *la sistematicità è una competenza emergente*. Anzi, secondo un autorevole membro del "gruppo PDP", Paul Smolensky, ogni tipo di competenza (ciò che può essere descritto come soddisfazione di regole concettuali) *emerge* da prestazioni di livello inferiore (dalla soddisfazione, cioè, di vincoli *sub*concettuali - come i pesi sulle connessioni) [Smolensky 1988, 107-8]. Secondo questa «teoria della competenza» un sistema connessionista

> fortemente distribuito può [...] esibire tutti i tipi di competenza comportamentale sistematica *senza* che quella competenza richieda spiegazioni in termini di composizionalità a livello concettuale [Clark 1989, 201].

Sono in molti a non ritenere esauriente questo tipo di risposta al problema della composizionalità e il dibattito innescato da Fodor e Pylyshyn è tuttora molto acceso (nella bibliografia di David Chalmers vi è un paragrafo specificamente dedicato a questo dibattito).

- *Il problema del livello di analisi*

Un altro problema affrontato esplicitamente dal "gruppo PDP" è quello relativo al livello di analisi. La questione è: *a quale livello si pone l'analisi connessionista dei fenomeni cognitivi?*

Questa domanda è stata avanzata da più parti[7] in un clima in cui era fortemente presente nei ricercatori della scienza cognitiva la celebre «teoria dei livelli di Marr». La teoria, sviluppata tra il 1976 e il 1982 (nei loro scritti sulla visione) da Reichardt, Poggio e il più noto David Marr, proponeva di suddividere l'analisi di ogni problema neurocomputazionale in una *gerarchia di livelli*. I "livelli di Marr" sono: il livello *computazionale* (analisi astratta del problema), il livello *algoritmico* (specificazione di una procedura formale per eseguire il compito), il livello *implementazionale* (costruzione di un dispositivo funzionante) [Rumelhart 1986, 170-2; Clark 1989, 31-2; Churchland 1992, 36-8]. Dove si collocano i modelli PDP?

---

[7]Rumelhart e McClelland ricordano ad esempio la critica di Donald Broadbent [Rumelhart 1986, nota 2 p.169].

La risposta "ufficiale" alla questione sollevata (a quale livello si pone l'analisi connessionista dei fenomeni cognitivi) è stata fornita da Smolensky in un articolo del 1988 (vedi Scheda 3). La risposta di Smolensky era in sintesi: secondo il paradigma connessionista la spiegazione formale completa della cognizione si trova a *livello subconcettuale* (!) [Smolensky 1988, 73]. Lungi dal risolvere il problema, l'articolo di Smolensky ha sollevato un polverone: decine di critiche sono piombate sui connessionisti e il numero delle «teorie dei livelli» è volato alle stelle. La questione è ancora aperta.

Scheda 3. *Sull'appropriato trattamento del connessionismo*

L'articolo di Paul Smolensky sull'approccio connessionista al problema della costruzione di modelli cognitivi è un lavoro ricco di spunti e idee interessanti, purtroppo rese in un linguaggio piuttosto ermetico e densamente occupato da pericolose "trappole" concettuali. Ad ogni passo ci si illude di poter dire: «Ho capito!» e subito si cade in una trappola: Smolensky è avaro di esempi e sembra contraddirsi innumerevoli volte. Tuttavia, alcune idee sono espresse con sufficiente chiarezza da poter essere riconosciute come vere e proprie sfide alla scienza cognitiva. Esse sono: l'idea del *paradigma subsimbolico*, l'idea dei *sistemi ibridi simbolici/subsimbolici* [Smolensky 1988, 88-9], l'idea dell'*emergenza simbolica* [Smolensky 1988, 106-9]. Per ragioni di spazio qui si può solo accennare alla prima di queste idee.

Il *paradigma subsimbolico* non è altro che l'approccio connessionista ai modelli cognitivi. Il concetto importante è però quello di "subsimbolo": entità con una duplice natura - semantica e sintattica. Dal punto di vista semantico, si tratta dei *costituenti dei simboli* (intesi in senso cognitivo classico); dal punto di vista sintattico, i subsimboli sono *operazioni computazionali numeriche*. Alla manipolazione simbolica del paradigma classico Smolensky oppone dunque l'idea dell'elaborazione subsimbolica e, poiché i subsimboli «*corrispondono all'attività delle unità di elaborazione individuali nelle reti connessioniste*» [Smolensky 1988, 62], l'elaborazione subsimbolica risulta essere null'altro che l'elaborazione distribuita in parallelo (PDP). I problemi vengono quando Smolensky afferma che il livello di descrizione fondamentale dei modelli PDP è quello subconcettuale, posto tra il livello neurale e quello concettuale [proposizione 11]. Non solo: a livello subconcettuale si trova anche la *spiegazione formale completa* della cognizione [Smolensky 1988, 73]. A questo punto ci aspetteremmo un illuminante esempio di analisi subconcettuale, ma le aspettative vengono deluse. Quanto di più "simile" ad un esempio si trova nel paragrafo 7.2, dove l'autore si propone di affrontare una questione fondamentale: la struttura costitutiva degli stati mentali. Smolensky fornisce la seguente definizione:

Uno stato mentale in un sistema subsimbolico è un *pattern* di attivazione con una struttura costitutiva che può essere analizzato sia a livello concettuale che a livello subconcettuale [Smolensky 1988, 96-7].

Un'analisi *a livello concettuale* è quella fornita da Pylyshyn nel 1984: la rappresentazione connessionista di *caffè* - dice ironicamente Pylyshyn - è ottenuta sottraendo alla rappresentazione *tazza con caffè* la rappresentazione *tazza senza caffè*. Smolensky afferma che, sempre a livello concettuale, tale sottrazione produrrebbe "in un certo senso" una rappresentazione connessionista di *caffè*, «*ma il caffè nel contesto della tazza*» [Smolensky 1988, 97]. Cioè, «*il* pattern *che rappresenta* caffè *nel contesto di* tazza *è diverso dal* pattern *che rappresenta* caffè *nel contesto di* caraffa, albero *o* uomo» [*ibid.*].

L'analisi concettuale, prosegue Smolensky, differisce da quella subconcettuale proprio per la diversa modalità di contestualizzazione [proposizione 23]: nel paradigma simbolico il contesto di un simbolo (per esempio il contesto del simbolo: *caffè*) si mostra attorno ad esso e consiste di altri simboli (*tazza*); nel paradigma subsimbolico il contesto di un simbolo si mostra *dentro* di esso e consiste di subsimboli (*liquido marrone con superficie piatta, liquido marrone a contatto con porcellana, contenitore rigido con maniglia, odore di bruciato,* ecc.)[8] [Smolensky 1988, 98]. Allora, riassumendo, uno stato mentale è -in un sistema connessionista- una struttura costituita da "simboli connessionisti" e i "simboli connessionisti" sono strutture costituite da subsimboli (tramite i quali si manifesta il contesto).

- *Il problema del controllo e la natura dei processi cognitivi*

Il più grave problema dei modelli PDP è, come già accennato, quello relativo alle facoltà cognitive "*di alto livello*": l'esecuzione di azioni pianificate, sequenzializzate e organizzate in strutture comportamentali coerenti. Questo problema si presenta sotto molte forme ed è al centro delle critiche al connessionismo più importanti: le reti imparano a seguire delle *regole* o si limitano a scoprire delle *regolarità*? Sono (pessimi) modelli dei fenomeni *cognitivi* o sono (discreti) modelli dei fenomeni *percettivi*?

E' interessante esaminare brevemente cosa rispondono i connessionisti, perché ci permette di capire qual è il loro reale atteggiamento nell'ambito della "febbre dell'oro" della mente. In due importanti paragrafi del testo del "gruppo PDP" [Rumelhart 1986, cap.6, §§3-4] il problema del controllo viene collegato esplicitamente alla concezione connessionista della natura del pensiero. E'

in altre parole venuto il momento di chiederci: *che natura hanno, secondo i connessionisti, i processi cognitivi?*

In generale, noi crediamo che i fenomeni cognitivi emergano dall'interazione di grandi insiemi di unità. Possiamo dunque considerare il livello di analisi simbolico come un'approssimazione del sistema soggiacente. In molti casi, queste approssimazioni si riveleranno utili; in altri, esse saranno sbagliate, e per comprendere il comportamento del sistema saremo costretti a passare al livello delle unità [Rumelhart 1986, 312].

In altre parole: i processi cognitivi sono proprietà emergenti delle reti neurali biologiche, cioè del sistema nervoso, e la relazione tra i modelli cognitivi classici e i modelli PDP è, per lo più, una questione di livelli di analisi. Questo significa che l'applicazione seriale e conscia delle regole (il problema del controllo) e la capacità stessa di risolvere problemi logici (i processi cognitivi) *sono tali*, secondo il paradigma connessionista, *solo dal punto di vista del cognitivismo classico* ("l'occhio della mente"):

Nel nostro precedente lavoro sulla percezione delle parole [...] e sull'apprendimento della morfologia dell'inglese [...] abbiamo dimostrato quanto è potente questo modo di impostare le cose. In quei casi siamo stati in grado di dimostrare come potesse facilmente *emergere* un'apparente applicazione di regole dalle interazioni tra unità semplici di elaborazione, senza dovere invocare nessuna applicazione di regole di livello superiore [Rumelhart 1986, 168].

L'idea di fondo è che la nostra capacità di risolvere problemi logici non è tanto basata sull'uso della logica, quanto sulla riduzione dei problemi che vogliamo risolvere a problemi che siamo capaci di risolvere. Gli esseri umani sembrano possedere *tre fondamentali capacità* che consentono loro di pervenire a conclusioni logiche senza essere logici [Rumelhart 1986, 297 - corsivo mio].

Queste tre capacità rappresentano il "nocciolo" della concezione connessionista della mente:

1. *Capacità di elaborazione distribuita in parallelo*;
2. *Capacità di costruire rappresentazioni interne*;
3. *Capacità di costruire rappresentazioni esterne*.

L'idea è la seguente:

1. *Il riconoscimento di configurazioni è la modalità computazionale di base.*

Il sistema umano di elaborazione delle informazioni va concepito fondamentalmente come una rete neurale, la quale *non calcola una soluzione* ma *si colloca in una soluzione* (rilassamento).

---

[8]In realtà, dice Smolensky, queste caratteristiche *non* sono veri e propri subsimboli, ma «*sono di livello sufficientemente basso da servire allo scopo*» [Smolensky 1988, 97].

Gli esseri umani sono capaci di «assestarsi» rapidamente in soluzioni di riconoscimento di configurazioni (cioè scoperta di regolarità) simili a quelle trovate dai modelli PDP. In altre parole, alla base della cognizione c'è la percezione: gli esseri umani *percepiscono* le soluzioni dei problemi (una volta ridotti ai minimi termini);

2. *Gli esseri umani sono abili nell'elaborare modelli del mondo.*

Questo è il passaggio-chiave dei processi cognitivi. E' la capacità di costruire *rappresentazioni interne* a consentire le «simulazioni mentali», indispensabili per la sopravvivenza di tutti gli organismi per i quali l'apprendimento svolge un ruolo cruciale;

3. *Gli esseri umani sono abili nel manipolare fisicamente il proprio ambiente.*

Per risolvere problemi, e quindi elaborare logica, matematica ed in generale produrre una cultura, è necessario ridurli ai minimi termini (in modo che le soluzioni possano essere percepite mediante meccanismi di tipo PDP). E per far questo diventa critica la capacità umana di creare *rappresentazioni esterne* del problema mediante la manipolazione dell'ambiente.

Consideriamo per esempio il problema di moltiplicare 343 × 822. *Manipolando l'ambiente* (capacità 3), ad esempio scrivendo su carta i due numeri in un particolare formato (rappresentazione esterna del problema), possiamo ridurre il problema ad una serie di operazioni ciascuna alla portata dell'elaborazione distribuita in parallelo del nostro cervello. Cioè possiamo *percepire* (capacità 1) che sotto 3 e 2 bisogna scrivere 6. E così via, continuando a modificare la rappresentazione esterna del problema, fino alla sua soluzione. Quest'esperienza viene *interiorizzata* (capacità 2) aggiornando la nostra rappresentazione interna del mondo. E' questo aspetto che consente (nei casi più semplici) di risolvere i problemi mentalmente.

Condividiamo la conclusione di Rumelhart e colleghi:

> Queste idee sono estremamente speculative e svincolate da qualsiasi particolare modello. A nostro giudizio, la loro utilità dipende dal fatto che suggeriscono in che modo i modelli di tipo PDP possono far fronte ad una classe di fenomeni per cui essi appaiono, a tutta prima, non del tutto appropriati - e cioè, essenzialmente, i fenomeni sequenziali e consci. In realtà, queste idee offrono delle nuove prospettive anche su questi fenomeni [Rumelhart 1986, 303].

Conclusione

Riteniamo che il connessionismo abbia un merito fondamentale: ci insegna che non si può inseguire la mente senza incontrare il cervello. I suoi limiti sono stati ampiamente mostrati e tuttavia l'idea dell'ispirazione neurale non ci sembra indebolita.

Non è possibile sapere cosa accadrà nei prossimi anni nell'ambito della scienza cognitiva. E' possibile che i contributi più interessanti *non* vengano dalla neurofisiologia, dal momento che difficilmente a quel livello si potrà trovare la mente. D'altra parte è impensabile qualsiasi ritorno al dualismo o ad altre filosofie svincolate dalla biologia.

Se dovessimo auspicare una direzione di ricerca, indicheremmo lo studio dei fenomeni emergenti. E' probabile che la mente si trovi in quei paraggi, ma se anche così non fosse il concetto di emergenza potrebbe risultare il punto di raccordo tra funzionalismo e connessionismo. Certamente, infatti, le proprietà emergenti delle attuali reti neurali sono troppo misere per fornirci indicazioni vere e proprie sulla mente, laddove il funzionalismo procede invece a grandi balzi. In un certo senso il rischio è che compiendo questi balzi il funzionalismo superi, senza accorgersene, la mente stessa.

Un'altra direzione che può fornire utili elementi di riflessione potrebbe essere l'endocrinologia. Attraverso lo studio delle basi biologiche delle emozioni, potremmo riuscire a discernere l'apparato emotivo della mente da quello puramente cognitivo e forse vedremmo con maggiore chiarezza quali sono gli aspetti davvero computazionali della mente. Potremmo anche scoprire che non è possibile isolare la cognizione, a nessun livello.

In questo caso tutti i nostri modelli computazionali andrebbero ripensati.

Bibliografia

Abu-Mostafa Y.S. - Psaltis D.
1987 "Il calcolatore ottico neuronico", in: **Lolli G.** (a cura di), *Mente e macchina*, «Le Scienze quaderni» n.66 (1992)

Churchland P.M.
1989 *La natura della mente e la struttura della scienza*, Il Mulino, Bologna 1992

Churchland P.S. - Sejnowski T.J.
1992 *Il cervello computazionale*, Il Mulino, Bologna 1995

Clark A.
1989 *Microcognizione*, Il Mulino, Bologna 1994

Corcoran E.
1991 "Calcolatori superveloci", in: **Lolli G.** (a cura di), *Mente e macchina*, «Le Scienze quaderni» n.66 (1992)

Fodor J.A.
1981 "Il problema mente-corpo", in: **Lolli G.** (a cura di), *Mente e macchina*, «Le Scienze quaderni» n.66 (1992)

Hillis W.D.
1987 "La «Connection Machine»", in: **Lolli G.** (a cura di), *Mente e macchina*, «Le Scienze quaderni» n.66 (1992)

Llinas R. - Pellionisz A.
1984 "La mente in quanto proprietà tensoriale dei circuiti cerebrali", in: **Piattelli-Palmarini M.** (a cura di), *Livelli di realtà*, Feltrinelli, Milano 1984

Marconi D.
1995 "Filosofia del linguaggio", in: **Rossi P.** (a cura di), *La filosofia*, UTET, Torino 1995, vol.I

Penrose R.
1989 *La mente nuova dell'imperatore*, Rizzoli, Milano 1992

Popper K.R. - Eccles J.C.
1977 *L'io e il suo cervello*, Armando, Roma 1992

Rumelhart D.E. - McClelland J.L.
1986 *PDP. Microstruttura dei processi cognitivi*, Il Mulino, Bologna 1991

Searle J.R.
1980 "Menti, cervelli e programmi", in: **Tonfoni G.** (a cura di), *Menti, cervelli e programmi. Un dibattito sull'intelligenza artificiale*, Clup-Clued, Milano 1984
1992 *La riscoperta della mente*, Bollati Boringhieri, Torino 1994

Smolensky P.
1988 "Il connessionismo", in: **Frixione M.** (a cura di), *Il Connessionismo tra simboli e neuroni*, Marietti, Genova 1992

Tank D.W. - Hopfield J.J.
1988 "Circuiti elettronici basati su modelli biologici", in: **Lolli G.** (a cura di), *Mente e macchina*, «Le Scienze quaderni» n.66 (1992)

Varela F.J. - Thompson E. - Rosch E.
1991 *La via di mezzo della conoscenza*, Feltrinelli, Milano 1992