

Sandro Nannini

Connessionismo e filosofia della mente

Abstract - The aim of this paper is to investigate the philosophical implications of the debate between the supporters of classical Artificial Intelligence (AI) (semantic networks, frames, production systems etc.) and the upholders of connectionism (particularly neural networks). However the analysis of that debate requires the three different ways to look at the AI itself to be distinguished.

1) The AI *simulates* by computers some cognitive activities of human beings, but it does not aim at *reproducing* the working of human minds. You can simulate by a computer for example the flight of a space capsule, but the simulated capsule does not indeed fly. In a similar way a 'thinking machine' can solve some problems like a human expert, but it does not really think because its inner working is probably different from the working of a human mind. This is the best meaning you can give the AI from the point of view of its technological development; but it is also an inadequate one from the point of view of its implications for the philosophy of mind.

2) The AI collaborates with logic in order to look for formal properties of those arguments that are intuitively correct. But logic has never been a good guide for psychology, and the AI according to this second meaning has nothing to do with cognitive sciences and the philosophy of mind.

3) The AI belongs to cognitive sciences on the ground prepared by functionalism, a philosophy of mind according to which mental states are similar to the functional states of a Turing machine (and therefore of a digital computer). The AI, if it is so meant, is the basis of the *computational theory of mind*, according to which being intelligent is only being programmed so as to be able to process information in a certain way independently of the material support of that operation. The computational theory of mind is a common presupposition both of classical AI and connectionism. Therefore I shall take it for granted in my analysis of the discussion between those two theories and shall not occupy myself with the criticism that J. Searle and others have addressed to it.

First of all, for it is necessary to specify at which level (or levels) of analysis that dispute has been managed, let me make clear what 'level of analysis' means in this context. The simplest way to do that is to refer to 'Marr's cascade' according to which, if you want to find a program to solve a certain problem, you must: 1) find in general terms a series of steps that lead on the wanted solution; 2) find for every step an algorithm able to execute it; 3) make sure that those algorithms can be implemented on the hardware of your computer.

The classical AI and the neural networks of connectionism differ surely at the third level (serial vs. parallel processing etc.). Do they differ however only at that level? Here the opinions are very different.

- The opponents of connectionism (like for example J.A. Fodor and Z.W. Pylyshyn) answer 'Yes' to that question. Furthermore, they think that the aim of cognitive sciences is to find out the functional architecture of human minds at the first and second level of Marr's cascade. According to them, minds are programs implemented on brains: cognitive psychologists are interested in mental software and not in the brain that implements it and which is studied by neuroscientists. Therefore, the opponents of connectionism think that, at the most, the neural networks offer to the AI an alternative implementation of classical algorithms, while they are insignificant for cognitive scientists and philosophers of mind.

The point of view of the upholders of connectionism is completely different. They see in connectionism the germ of a new paradigm for cognitive sciences. Within this common horizon I shall reconstruct three different theories relying (but rather freely) on the positions expressed by Paul and Patricia Churchland, P. Smolensky and A. Clark.

The first theory (the most radical one) maintains three main items:

1) Any separation between psychology and neurosciences is a tailing of an old

metaphysical dualism between mind and body. Furthermore, the neural networks seem to have performances and failings very similar to those typical of human minds. Therefore they make it easier to imagine a plausible answer (may be for the first time) to an old question: 'How on earth can a body think?'

2) It is biologically implausible that human minds have a structure which is fundamentally linguistic. It is particularly implausible that sensorimotor coordinations are operated by means of a *manipulation of symbols* (as a robot driven by a digital computer does) because they are the first intelligent performance made perfectly well by non speaking animals. Instead of projecting back 'linguistic' abilities like the manipulation of symbols on the brains of animals, it is better to take the other way round and imagine how animal brains that were able to operate sensorimotor coordinations in a simple way became by evolution human brains able also to speak. The neural networks offer a good hypothesis about that simple and non 'linguistic' way of operating sensorimotor coordinations.

3) Connectionism is part of a broader scientific revolution that is leading psychologists to give up the folk psychology and replace it by a new theory of human behaviour grounded on neurosciences. Therefore connectionism joins from a philosophical point of view the so-called eliminativistic materialism upheld by R. Rorty, S. Stich, the Churchlands and others.

- The other two ways of understanding connectionism differ from the first one because they do not accept the eliminativistic materialism.

Smolensky thinks first of all that the descriptions and explanations of mental facts given by the folk psychology and the classical AI are not wrong but approximate and idealized. Secondly, he thinks that neural networks (notwithstanding their name) describe and explain the working of cognitive system at a level intermediate between the physical level of neurons and the manipulation of symbols according to syntactic rules. For example in a discussion with Pylyshyn, Smolensky shows that you can give three distributed representations of 'cup with coffee', 'cup without coffee' and 'coffee' so as that if you say that the first representation is the result of a combination of the other two you are not wrong. Your description of the working of that network is however approximate. You describe the network in general terms as if it had at its disposal a concept of coffee independent of context and could combine it with 'cup' or for example with 'can'. But in fact it is not exactly so. If you look at the microfeatures that form the representation of coffee given by the network, like for example 'brown liquid contacting porcelain', you see that the coffee represented by the network is the coffee in the context of a cup and not the general (and freely combinable) concept of coffee used at the symbolic level of analysis. Describing the working of a cognitive system, the relationship between the symbolic and the subsymbolic level is the same as between classical and quantum mechanics. The rules followed by cognitive systems at the symbolic level are idealizations to which the resultant of the working of a myriad of processors at the subsymbolic level asymptotically approximates if the processing time is unlimited and the input without noise.

According to Clark, Smolensky's theory is still too favourable to eliminativistic materialism because Smolensky thinks that only at the subsymbolic level can one explain exactly the working of cognitive systems whereas symbolic level explanations are only approximate. Instead, Clark thinks that only at the symbolic level can one show some general and high-order features of neural networks that emerge when these acquire a certain degree of complexity.

Furthermore Clark thinks that the symbolic level is not only an external point of view over the neural networks and the minds of other people; it is also the point of view adopted by human brains in order to selfmonitor their own inner working and improving in such a way their performances. Of course this selfmonitoring is nothing else than consciousness (understood like a form of endopsychical perception). This hypothesis is still only a untested suggestion; however it shows clearly enough that Clark does not oppose the classical AI and connectionism but see them as two points of view that can collaborate in explaining the structure and working of human minds.

L'intelligenza artificiale classica (reti semantiche, *frame* lato, e S. Papert e M.C. Minsky (1969), dall'altro, intorno alle capacità del *perceptron* con un'architettura meno semplice, l'invenzione di nuovi algoritmi per il loro addestramento ed il ricorso al modello PDP (*Parallel Distributed Processing*), il connessionismo ha conosciuto un grande rilancio (Rumelhart e McClelland 1986). Le reti neurali riscuotono oggi un'attenzione sempre maggiore da parte di *computer scientists*, ingegneri, matematici, fisici, psicologi, linguisti e filosofi della mente; ma non sono mancati neppure studiosi che hanno manifestato totale scetticismo nei loro confronti e hanno ribadito la superiorità dell'intelligenza artificiale classica, mentre altri ancora hanno espresso l'opinione che il contrasto tra questi due indirizzi sia stato drammatizzato.^[1]

Ricostruire l'intera *querelle* dal punto di vista tecnico sarebbe molto lungo, difficile ed esorbiterebbe dalle mie competenze. Pongo perciò del tutto in parentesi questo aspetto del problema e do per sufficientemente noto che cosa siano e come funzionino le reti neurali. Il punto sul quale intendo soffermarmi concerne le implicazioni che esse hanno per la filosofia della mente. Ci richiede tuttavia alcune considerazioni preliminari sul modo d'intendere l'intelligenza artificiale in generale. Si danno, al riguardo, tre alternative principali.

1) Le macchine, per essere capaci di simulare alcune di tali attività giudicate intelligenti, ossia siano capaci di trovare, per ogni input appartenente ad un certo repertorio, l'output corretto corrispondente, indipendentemente dal fatto che il metodo adottato dalla macchina per risolvere il problema assegnatole assomigli o meno a quello che s'ipotizza venga impiegato (coscientemente o no) dagli uomini, gli effetti della selezione naturale, l'andamento demografico, l'evoluzione di un sistema economico ecc. Perché non tentare di simulare la riproduzione di ciò che viene simulato e neppure alcuna somiglianza sostanziale con esso. Così come le navicelle spaziali simulate al computer non volano, allo stesso modo le macchine pensanti a rigore non pensano. O meglio, se vengono forniti loro certi dati, esse giungono, in ambiti delimitati, a conclusioni simili a quelle alle quali giungerebbero degli esperti umani. Ma il modo nel quale operano le macchine (ossia nel quale elaborano l'informazione) può darsi che nulla abbia a che vedere con il funzionamento effettivo delle menti umane. E' questo il modo d'intendere l'intelligenza artificiale più

2) L'intelligenza artificiale può essere usata anche come una sorta di laboratorio al servizio della logica, o meglio al servizio della ricerca delle proprietà formali dei ragionamenti intuitivamente corretti. Si pensi, ad esempio, all'impulso reciproco che si sono date le ricerche sull'ereditarietà (con eccezioni), nelle reti semantiche e nei *frame*, di certe caratteristiche da concetti generali a

concetti particolari^[2] e le ricerche sulle logiche non monotoniche (Frixione 1994, pp. 151-174). E' questo un modo d'intendere l'intelligenza artificiale che ha evidenti legami con la logica e la filosofia in generale, ma non propriamente con le scienze cognitive e la filosofia

3) L'intelligenza artificiale pu~ essere considerata come una parte della scienza cognitiva entro la prospettiva filosofica del funzionalismo, secondo il quale intercorre un'analogia tra l'*hardware* e il *software* di un computer, da un lato, ed il cervello e la mente dell'uomo, dall'altro. Gli stati mentali sono equiparabili a stati funzionali implementati sul *wetware* cerebrale. Gli studi d'intelligenza artificiale in questa terza accezione - che J.R. Searle (1980), per rifiutarla, ha chiamata 'forte' - offrono in modo diretto allo psicologo cognitivo ipotesi e controprove per la ricostruzione dell'architettura funzionale della mente umana e vengono a costituire, sotto l'ombrello filosofico del funzionalismo, il fondamento di quella concezione che va sotto il nome di teoria computazionale della mente, secondo la quale un sistema fisico, sia esso il cervello umano o un cuitivamente, cognitive (ad esempio, giocare a scacchi). Questo terzo modo d'intendere l'intelligenza artificiale, in quanto fond. Le differenze riguardano solo le modalit~ di tale processamento.

9), J.R. Searle (1980) e D.R. Hofstadter (1985). Ma Dreyfus e Hofstadter hanno criticato la sua versione classica prima del rilancio del connessionismo e difeso idee in parte con esso convergenti.^[3] il prodotto di poteri causali che appartengono solo alla materia cerebrale. Un supporto fisico diverso, quale che sia il suo stato funzionale, non pu~ riprodurle. La sua critica, contestando la possibilit~ stessa di un'intelligenza *artificiale* che riproduca (e non solo simuli) quella umana, colpisce perci~ alla radice l'analogia tra menti e stati funzionali del computeligenza artificiale classica quanto contro il connessionismo (se queste teorie pretendono di dare vita ad un'intelligenza artifiuello di mettere in luce le implicazioni filosofiche delle differenze che intercorrono tra intelligenza artificiale classica e connessionismo, e non di analizzare le critiche ai loro presupposti comuni.

Il confronto tra queste due prospettive risulta per~ sfuocato se non viene precisato che la teoria computazionale della mente pu~ muoversi a differenti livelli d'analisi. Occorre perci~ chiarire questo concetto (e quello connesso di macchina virtuale).

di D. Marr (1982). Questi ha notato che un programmatore, quando cerchi di scrivere un programma capace di eseguire operazioni cognitive, affronta il problema *top-down* a tre livelli diversi.

Anzitutto egli cerca di chiarire, in termini generali e mediante concetti appartenenti allo stesso ambito (o *task domain* In modo analogo uno scolaro delle elementari, posto di fronte ad un 'problema', deve prima di tutto individuare quali operazioni debba eseguire fra i dati a sua disposizione: ad esempio, deve individuare una regola del tipo 'Per calcolare il guadagno devo

sottrarre le spese dal ricavo'.

In secondo luogo il programmatore deve inventare, per ciascuno di questi passi, un algoritmo capace di eseguirlo (e scrivere quest'ultimo in un linguaggio disponibile; ad esempio in LISP). Similmente lo scolaro deve saper eseguire le singole operazioni aritmetiche individuate al livello precedente (avvalendosi dell'opportuno incolonnamento di numeri scritti con il sistema decimale ecc.).

ggio da lui adottato al secondo livello (ad esempio il LISP). In altre parole il computer ha un'architettura interna tale che, di leggi naturali un altro evento fisico decodificabile come la risposta cercata. Ad esempio, quando uso una semplice calcolatrice di leggi naturali e della struttura dei circuiti della macchina, la comparsa sullo schermo (altro evento puramente fisico) di una figura a forma di "5". La macchina reale, fatta solo di circuiti, di tasti ecc., non sa nulla delle regole dell'aritmetica, ma, se i suoi input ed output vengono opportunamente codificati e decodificati, funziona come se le conoscesse: o meglio, implemevello) dallo scolaro per eseguire la sottrazione (la cui necessit  per risolvere il problema era stata individuata al primo livello) sia implementato nel suo cervello (al terzo livello) mediante l'accensione di determinati neuroni.

I tre livelli di Marr (computazionale, algoritmico e implementativo) possono sdoppiarsi ulteriormente e dare vita ad un'analisi o codice interpretativo, il funzionamento di una macchina virtuale descritta ad un livello di maggiore generalit ; quest'ultima " via. La stessa macchina pu  essere descritta al primo livello d'analisi, ad esempio, come un giocatore di scacchi, al secondo come un processore di liste (in linguaggio LISP), al terzo come un computer digitale.

Ora, ci si pu  chiedere, a quale livello (o a quali livelli) della cascata di Marr si colloca il contrasto tra intelligenza artificiale classica e connessionismo? Una differenza evidente si trova sicuramente al terzo livello. Le reti neurali, anche se vengono spesso simulate su computer digitali seriali, sono tuttavia delle macchine (reali o virtuali) che funzionano in parallelo. Le differenti prestazioni dei due tipi di processamento sono ben note. In particolare, i computer digitali sono dei `sistemi simbolici fisici', nel senso che A. Newell e H.A. Simon (1976) hanno dato a questa espressione, e perci  implementano con facilit  delle macchine virtuali capaci di trarre inferenze; ossia di trasformare, secondo regole logico-linguistiche, delle stringhe di *simboli* in altre stringhe anch'esse formate da simboli appartenenti al medesimo repertorio dei precedenti. Il processamento avviene secondo regole sintattiche che la macchina applica ai simboli solo per la loro forma; ma, dal punto di vista dell'utente esterno, to) che si mantiene costante non solo nell'input e nell'output, ma anche in tutti i passaggi intermedi. Le reti neurali, viceversa, elaborano i loro dati mediante un processo che non ha nulla in comune con un'inferenza guidata da regole logico-linguistiche; sono, in compenso, particolarmente adatte

all'autoprogrammazione (o 'apprendimento') mediante addestramento e riescono bene nel riconoscimento pressoché istantaneo di *pattern* complessi.

L'elenco delle differenze di struttura, funzionamento e prestazioni potrebbe allungarsi molto. Ma la domanda rilevante, per chi 'altra: i due approcci divergono solo a livello implementativo oppure anche a quello algoritmico e computazionale? Qui i pareri sono decisamente discordi.

I fautori dell'intelligenza artificiale classica hanno sostenuto che le reti neurali, ammesso che siano di qualche utilità, rappresentano una novità solo a livello implementativo (Fodor e Pylyshyn 1988). Quanto questo giudizio sia critico verso il connesse all'intelligenza artificiale rilevanti per la scienza cognitiva si collocano al primo e secondo livello della cascata di Marr. L'ordine, ispirandosi talvolta ai programmi dell'intelligenza artificiale, l'architettura funzionale della mente umana, senza preoccuparsi di come essa sia implementata sul cervello. Ovviamente egli è per scontato che in qualche modo essa lo sia; ma scoprire l'analisi (1) e (2), da un lato, e (3), dall'altro, della cascata di Marr - ben poco hanno a che fare con le scienze cognitive e la filosofia della mente.

tenere che esso riveste scarsissimo interesse per le scienze cognitive e la filosofia della mente. Anche ammesso che sia possibile implementare mediante reti neurali i singoli passi di un algoritmo capace di calcolare una funzione cognitiva, ciò non tocca la natura dell'algoritmo stesso, che, in quanto macchina virtuale intelligente, non può non combinare simboli e trasformare loro stringhe secondo regole sintattiche e d'inferenza, perché giustappunto (e unicamente) in queste capacità di processamento dell'informazione mediante regole di tipo linguistico consiste l'intelligenza. Il fatto che le reti neurali non siano dei 'sistemi simbolici fisici' e che il loro funzionamento, lungi dall'essere retto da regole di tipo linguistico, sia governato in modo statistico ed olistico da equazioni differenziali simili a quelle studiate dalla termodinamica, non tocca la natura 'simbolica' delle macchine virtuali che esse implementano, quando queste ultime siano dei sistemi intelligenti. Essi, quali che siano le modalità di funzionamento del loro supporto fisico, non possono non manipolare simboli secondo regole sintattiche, perché solo la manipolazione di simboli offre, mediante il 'principio di composizionalità' del linguaggio, quella capacità di creare e riconoscere com

tare le singoli parti funzionali di una macchina virtuale classica.

Riassumendo, si può dire che la posizione dei critici del connessionismo si articola in due punti: 1) La struttura profonda del ente riferirsi a questa tesi dandole il nome di 'razionalismo'. 2) La mente umana, in quanto macchina virtuale implementata sulle scienze.

mutamento di paradigma delle scienze cognitive. Sono concepibili tuttavia, entro questo orizzonte comune, posizioni piuttosto diverse. Ne metter

in luce tre, appoggiandomi (ma in modo piuttosto libero) agli studi, rispettivamente,, di Paul e Patricia Churchland (1992 e 1986), di P. Smolensky (1991 e 1992) e di A. Clark (1994).

o un alleato potente nella difficoltà che incontravano i materialisti a, non dico spiegare, ma anche solo lontanamente immaginare l'eseguire calcoli, ma per un altro anche terribilmente meno duttili ed intuitive - presentano ancora differenze tali rispetto all'uomo che non sembrano aver risposto esaurientemente alla domanda se e come la materia possa davvero pensare. Le reti neurali, positive e negative, analoghe a quelle delle menti umane^[4] - cominciano oggi ad offrire una risposta plausibile, ancorché incompleta e non definitivamente provata, alla domanda 'Come può il cervello umano implementare una macchina virtuale intelligente?'. Su questa base il neuroscienziato, lo psicologo cognitivo ed il *computer scientist* possono utilmente collaborare.

Infatti il risultato di un'evoluzione che ha preso le mosse da cervelli di animali non parlanti, la cui funzione primaria era il un'architettura funzionale propria di una facoltà, il linguaggio, comparsa al termine e non all'inizio di tale processo evolutivo. Se vogliamo spiegare in modo biologicamente attendibile lo sviluppo nella specie umana di quel complesso di capacità che (con insufficiente precisione) chiamiamo intelligenza, dobbiamo non proiettare all'indietro sul coordinamento sensomotorio l'architettura funzionale tipica del linguaggio (come fa invece un robot guidato da un computer digitale), bensì all'inverso spiegare come dei sistemi capaci di operare il coordinamento sensomotorio grazie ad un metodo molto semplice e facilmente implementabile su una rete di neuroni e di sinapsi, possano avere acquisito la facoltà di manipolare simboli. Ora, le reti neurali, insieme ad altri sistemi che s'ispirano anch'essi alla struttura del cervello (Churchland 1992, pp. 57-95), sembrano essere particolarmente adatte a realizzare in modo semplice (e al tempo stesso veloce) il coordinamento sensomotorio, perché possono assolvere a questo compito mediante un'elaborazione di dati che richiede non già una lunga manipolazione seriale di simboli secondo regole sintattiche, bensì la trasformazione del vettore m-dimensionale di input nel vettore n-dimensionale di output mediante un processo eseguito in par delle connessioni, secondo un'equazione differenziale che ne determina i valori rispetto al tempo.^[5]

folk-psychology con tutti i suoi concetti (credenza, desiderio, intenzione ecc.) ed alla sua sostituzione con una scienza cognitiva costruita a partire dalle neuroscienze. La *folk-psychology* e i desideri cos' come non esiste il flogisto.

co nel materialismo eliminativistico di R. Rorty (1965), di S. Stich (1994) e degli stessi Churchland, secondo i quali i materialisti devono non ridurre gli stati mentali della *folk-psychology* a stati fisici del sistema nervoso centrale, come

cercano di fare invece i sostenitori della *Mind-Brain Identity Theory*,^[6] bensì mostrare che i concetti che esprimono tali stati mentali sono prescientifici e possono essere sostituiti vantaggiosamente con altri attinti dalle neuroscienze.

Smolensky (1991 e 1992) ritiene, in primo luogo, che le descrizioni dei fenomeni mentali offerte dalla *folk-psychology* a livello concettuale, ossia ottenute usando termini psicologici che si ritrovano nel linguaggio comune (un metodo al quale anche l'intelligenza artificiale classica s'ispira largamente), non siano completamente errate. Esse colgono, a un livello molto generale d'analisi, delle macrocaratteristiche che, *con un certo grado d'approssimazione*, emergono dal funzionamento di reti neurali complesse. In secondo luogo egli pensa che tali reti, funzionando ad un livello intermedio *subconcettuale* ed essendo dei sistemi cognitivi *subsimbolici*, consentano di gettare un ponte tra la descrizione dei fenomeni a livello, rispettivamente, concettuale e cerebrale. Con ciò Smolensky si allinea a Fodor e Pylyshyn quanto a Churchland; ma non così alto come quello generico e approssimato della *folk-psychology* o dei programmi dell'intelligenza artificiale classica. I sistemi cognitivi subsimbolici rendono evidente come una macchina (reale o virtuale) che presenti un'architettura funzionale connessionistica, quando superi una certa soglia di complessità e ricorra a rappresentazioni distribuite,^[7] acquisisce automaticamente delle caratteristiche che la rendono approssimativamente simile ad una macchina virtuale (di livello superiore) capace di manipolare simboli seguendo regole rigide.

Chiariamo questo punto con un esempio. Nel corso di un convegno Z.W. Pylyshyn aveva rivolto a Rumelhart ed altri la seguente obiezione^[8]

Smolensky (1991, p. 291 sgg.) mostra, mediante un 'esempio-giocattolo', che le cose non stanno precisamente così: consistono in un *pattern* di valori d'attivazione (per semplicità solo 1 o 0) di dieci nodi di una rete. Ciascuno di questi nodi rappresenta, singolarmente, la presenza (1) o assenza (0) di una microcaratteristica (*microfeature*) differenti.

	Tazzina		
Contenitore verticale	1	1	0
Liquido caldo	1	0	1
Vetro a contatto con il legno	0	0	0
Superficie curva di porcellana	1	1	0
Odore di bruciato	1	0	1
Liquido marrone a contatto con la porcellana	1	0	1
Superficie curva di porcellana	1	1	0
Oggetto oblungo d'argento	0	0	0

Manico adatto alle dimensioni d'un dito	1	1	0
Liquido marrone con lati curvi e fondo	1	0	1

Come si vede, in questo esempio-giocattolo i nodi della rete sono etichettati: ciascuno rappresenta una microcaratteristica esprimibile mediante una stringa non troppo lunga di parole del linguaggio comune. In realtà molte reti neurali, qualora aggiustino i loro pesi automaticamente mediante addestramento, non sono cos' "razionali" (come vuole testimoniare, nell'esempio stesso, l' "inutile" ripetizione della microcaratteristica "superficie curva e un valore semantico "comprensibile" a certi *pattern* attribuire ad ogni singolo nodo un significato esprimibile mediante una parola o un numero non troppo grande di parole. Il modo nel quale la rete classifica e organizza i suoi input, mediante i *pattern*

Ci non vuol dire però che le rappresentazioni distribuite delle reti neurali siano necessariamente prive di una struttura comp' espressa da *pattern* loro profilo comune, questa loro approssimativa coincidenza, che viene colta, a livello simbolico, dal concetto corrente di caffè basata sulla semplice combinazione di simboli. Nella descrizione del funzionamento di un sistema cognitivo - conclude Smolensk microfisica quantistica. Le regole che seguono i sistemi cognitivi a livello simbolico sono delle idealizzazioni alle quali si approssima asintoticamente la risultante del lavoro di una miriade di processori a livello subsimbolico, quando il tempo di elaborazione sia illimitato e i dati disponibili privi di rumore.

anzitutto contesta a Smolensky di lasciare ancora troppo spazio all'eliminativismo, allorché afferma che le analisi a livello simbolico del funzionamento dei sistemi intelligenti sono delle mere approssimazioni. Questa tesi lascia aperta la porta a chi voasse (ossia quello cerebrale-implementativo) sia completamente accettabile. In realtà - osserva Clark (1994, p. 265), richiamandosi esplicitamente alla *intentional stance* e la previsione del comportamento dei sistemi intelligenti risulterebbero cos' complesse da essere praticamente impossibili. Se cessassimo di attribuire ai nostri simili desideri e credenze e pretendessimo apaci di vivere con loro. In modo analogo, può darsi benissimo che i *frame* o i "sistemi di produzione" dell'intelligenza artificiale classica mettano in luce alcune, almeno, delle proprietà di funzioname tra i valori di una miriade di nodi e di connessioni, sarebbe praticamente impossibile cogliere.

In secondo luogo Clark avanza l'ipotesi (per ora, certo, non provata) che il cervello degli esseri umani sia capace, automonitorandosi, di mettere in luce e riutilizzare quelle caratteristiche che, a livello simbolico d'analisi, emergono tra i *pattern* d'attivazione interni delle reti neurali che, almeno in buona parte, lo

costituiscono. Secondo tale ipotesi il livello simbolico di un'autoanalisi che ogni singolo cervello utilizza per migliorare le proprie prestazioni. Questa funzione di autoanalisi

E' per ora prematuro giudicare se queste suggestioni di Clark, anch'esse in ultima analisi ispirate a Dennett (1993), abbiano un'intelligenza artificiale classica o quello del connessionismo, si sforza di mostrare come questi due approcci alle scienze cognitive possano cooperare ad una migliore comprensione della mente umana.

Note

[1] 1991) e Ramsey, Stich e Rumelhart (1991).

[2]rchŽ Fido, per un difetto genetico o in conseguenza di un incidente, potrebbe averne, poniamo il caso, tre.

[3] P. Smolensky (1992, p. 245), uno degli autori del rilancio del connessionismo, ha addirittura riconosciuto in Hofstadter uno dei suoi principali ispiratori.

[4] Si pensi, ad esempio, alla *graceful degradation*, alla *content-addressed memory*, alla generalizzazione automatica ecc. Su questo punto sia consentito rinviare a Nannini (1995).

[5] m nodi di input costituiscono le coordinate di un certo punto in uno spazio a m dimensioni, che possiamo chiamare lo 'spazio del problema'. Quando la rete ha la struttura adatta ed ha trovato mediante addestramento (o gli sono stati forniti direttamente dal ricercatore) i pesi giusti sulle connessioni, ogni input genera in modo automatico il suo output corretto, consistente nei valori di attivazione di n nodi. Questi valori costituiscono le coordinate, in uno spazio a n dimensioni che possiamo chiamare lo 'spazio delle soluzioni', del punto che rappresenta la soluzione cercata (o di un punto ad esso sufficientemente vicino). Le reti neurali sono perciò una generalizzazione dei metodi indicati da Paul Churchland per operare il coordinamento sensomotorio mediante la trasformazione di coordinate.

[6] Per un quadro d'insieme delle varie versioni contemporanee della teoria dell'identità fra mente e cervello e le relative obiezioni, cfr. Macdonald (1989).

[7] d'attivazione di molteplici nodi e non nel valore d'attivazione di uno solo. Per un esempio, molto semplificato, di rappresentazioni distribuite cfr. *infra*

[8] Laughlin (1991).

Bibliografia

Churchland, Patricia: 1986, *Neurophilosophy: Toward a Unified Understanding of the Mind-Brain*, MIT Press, Cambridge, MA.

Churchland, Paul: 1992, *La natura della mente e la struttura della scienza. Una prospettiva neurocomputazionale*, trad. it., Il Mulino, Bologna.

Clark, A.: 1994, *Microcognizione. Filosofia, scienza cognitiva e reti neurali*, trad. it., Il Mulino, Bologna.

Dennett, D.: 1987, *The Intentional Stance*, MIT Press, Cambridge, MA.

Dennett, D.: 1993, *Coscienza*, trad. it., Rizzoli, Milano.

Dreyfus, H.: 1979, *What Computers Can't Do*, 2nd edn., Harper & Row, New York.

Fodor, J.A. and Z.W. Pylyshyn: 1988, 'Connectionism and Cognitive Architecture: A Critical Analysis', *Cognition* 28, 3-71.

- Fodor, J.A. and B.P. McLaughlin: 1991, 'Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work', in T. Horgan and J. Tienson (eds.), *Connectionism and the Philosophy of Mind*, Kluwer Academic Publishers, Dordrecht, Boston, London, pp. 331-354.
- Frixione, M.: 1994, *Logica, significato e intelligenza artificiale*, Angeli, Milano.
- Hofstadter, D.R.: 1985, 'Waking Up from the Boolean Dream, Or Subcognition as Computation', in Id., *Metamagical Themas: Questing for the Essence of Mind and Pattern*, Viking, New York.
- Horgan, T. and J. Tienson (eds.): 1991, *Connectionism and the Philosophy of Mind*, Kluwer Academic Publishing Company, Dordrecht, Boston, London.
- Macdonald C.: 1989, *Mind-Body Identity Theories*, Routledge, London, New York.
- Marr, D.: 1982, *Vision*, MIT Press, Cambridge, MA.
- Minsky, M.C. and S. Papert: 1969, *Perceptrons. An Introduction to Computational Geometry*, MIT Press, Cambridge, MA.
- Nannini, S.: 1995, 'Sistemi simbolici e reti neurali nella filosofia della mente', *L'immagine riflessa*, di prossima pubblicazione.
- Newell, A. and H.A. Simon: 1976, 'Computer Science as Empirical Inquiry: Symbols and Search', ristampato in M.A. Boden (ed.), *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990, pp.105-132.
- Ramsey, W., S.P. Stich, and D.E. Rumelhart (eds.): *Philosophy and Connectionist Theory*, Lawrence Erlbaum Associates Publishers, Hillsdale, N.J., Hove, London
- Rorty, R.: 1965, 'Mind-Body Identity, Privacy, and Categories', *Review of Metaphysics* **19**, 24-54.
- Rosenblatt, F.: 1962, *Principles of Neurodynamics*, Spartan, New York.
- Rumelhart, D.E. and J.L. McClelland: 1986, *Parallel Distributed Processing*, MIT Press, Cambridge, MA.
- Searle, J.R.: 1980, 'Minds, Brains, and Programs', ristampato in M.A. Boden (ed.), *The Philosophy of Artificial Intelligence*, Oxford University Press, Oxford, 1990, pp. 67-88.
- Smolensky, P.: 1991, 'The Constituent Structure of Connectionist Mental States: A Reply to Fodor and Pylyshyn' in T. Horgan and J. Tienson (eds.), *Connectionism and the Philosophy of Mind*, Kluwer Academic Publishing Company, Dordrecht, Boston, London, pp. 281-306.
- Smolensky, P.: 1992, *Il connessionismo tra simboli e neuroni*, trad. it., Marietti, Genova.
- Stich, S.: 1994, *Dalla psicologia del senso comune alla scienza cognitiva*, trad. it., Il Mulino, Bologna.