Università degli Studi di Catania
Facoltà di Lettere e Filosofia
Laurea in Filosofia
Anno Accademico 2004/2005

DAVIDE DELL'OMBRA

Neuroni e circuiti.
L'Intelligenza artificiale tra filosofia e scienza

Relatore
Prof. Sara Condorelli

SOMMARIO

Introduzione p. 3

Il rapporto tra filosofia e scienza nell'ambito dell'intelligenza artificiale

Capitolo 1.

Il problema mente-corpo in filosofia p. 11

La problematica collaborazione tra mentale e fisico nell'ottica filosofica

Capitolo 2.

Storia dell'Intelligenza Artificiale e teorie della mente p. 19

Nascita e sviluppi del progetto nei suoi legami indissolubili con la filosofia della mente

Capitolo 3.

Della realtà ultima p. 60

La prospettiva che la realtà ultima della mente sia la realtà ultima

Riferimenti bibliografici p. 74

«La verità è importante; ma anche l'importanza lo è»

H. Putnam[1]

«Come meravigliarsi se i filosofi, riflettendo sul problema della mente senza questo bagaglio di conoscenze, sono stati tentati di ipotizzare nuove entità, i fisici nuovi e misteriosi campi materiali, mentre quanti confidano nell'immortalità continuano a ipotizzare spiriti eterni?»

G. M. Edelman[2]

Introduzione

Sono tante le cose – verità, fatti, nature – che ignoriamo. Fra queste vi è la nostra stessa intelligenza. Ma i problemi ancora aperti nell'ambito dell'intelligenza naturale non hanno impedito l'affermarsi del *progetto* di Intelligenza Artificiale. Una questione che qui vorrei collegare a quella dell'identità stessa della filosofia oggi: si parla della sua fine, della dissoluzione del campo d'indagine della filosofia, invaso da altri saperi del tutto nuovi. La filosofia è troppo «tecnica»? Ha invaso indebitamente il campo specifico della scienza? Quale è l'ambito dei saperi e della realtà di cui la filosofia si deve occupare? Nemmeno i filosofi sanno dare risposte univoche a tali questioni, soprattutto riguardo l'oggetto d'indagine del loro stesso lavoro, in quanto pressoché ogni buon filosofo delinea sempre un proprio specifico campo d'indagine nell'introdurre il proprio pensiero, o addirittura propone un nuovo modo di concepire i confini della filosofia criticando i precedenti. L'intelligenza artificiale è un oggetto di studio che, grazie alla compenetrazione di teoria e pratica, incrocia trasversalmente le discipline anche se esse non hanno delineato precisamente i loro confini reciproci. In questo ambito la filosofia gioca alla pari, possiamo dire, con la scienza, intendendo questa come quella universalità di studi che tenta la spiegazione del reale in maniera alternativa rispetto all'approccio filosofico. Iniziamo così a distinguere scienza e filosofia, in via del tutto propedeutica.

Hilary Putnam, da attento filosofo della scienza quale è, nota che non si può assolutamente parlare di «fine della filosofia» giacché la tecnicità e la prossimità allo stile scientifico nelle argomentazioni (o, in alcuni casi filosofici, dimostrazioni vere e proprie) di determinate in-

[1] H. Putnam (1987), p. 333.

[2] G. M. Edelman (1993b), p. 230.

dagini prettamente filosofiche non è da considerarsi come condizione negativa bensì «salutare» della filosofia. Riporto qui le sue parole al riguardo, davvero molto efficaci: «Se servisse un'ulteriore prova dell'attuale stato di buona salute della filosofia, basterebbe citare le torme di intellettuali che lamentano l'eccessivo "tecnicismo" della filosofia odierna, la sua "rinuncia" a ogni interesse per i problemi "reali", e così via. Lamentele del genere, infatti, si sono sempre levate proprio nei momenti in cui la filosofia era importante e vitale! Aristofane trovava Socrate vacuo e tecnico; Berkeley fu giudicato ridicolo dall'opinione corrente finché Hume e Kant non riconobbero l'importanza della sua sfida; Hume e Kant vennero a loro volta ridicolizzati e fraintesi (Kant nientemeno che da un personaggio come Goethe, il quale vedeva in lui un buon cervello prostituitosi al servizio della Chiesa). Il guaio è che la buona filosofia è ed è sempre stata *ostica* e che è più facile imparare i nomi di quattro filosofi che leggere i loro libri. Coloro che oggi trovano la filosofia troppo "tecnica", sono gli stessi che non avrebbero trovato il tempo o l'interesse per seguire le lunghe catene del ragionamento socratico o per leggere una delle *Critiche*»[3].

Non mancano, per questo motivo, le critiche alla filosofia anche da parte dei filosofi stessi, come nel caso di Daniel Dennett. Il filosofo incentra la questione sull'«affetto» che molti filosofi provano nei confronti delle loro teorie; difatti Dennett afferma che «i filosofi hanno nei confronti delle loro teorie, lo stesso rapporto che alcuni genitori hanno con i loro figli: vorrebbero non abbandonarli mai, impedendo loro di crescere»[4]. Aggiungo che certo non è il caso, tra molti altri, del citato Putnam, infatti «lungi dal meritare l'ironia che gli viene di frequente riservata, testimonia solo la sua onestà intellettuale e la serietà con la quale, al di sopra di un astratto ideale di coerenza, egli persegue quella ricerca della verità che dovrebbe essere propria di ogni filosofo e scienziato»[5]. Ribadendo la propria idea, Dennett dice di immaginare la scelta di un filosofo rispetto ad uno scienziato se posto di fronte ad alternative quali la stesura di un testo la cui «soluzione sarà esauriente, definitiva e assoluta, al punto che la comunità scientifica e intellettuale, ma anche la gente comune, ne parlerà per mesi,

[3] H. Putnam, *Recensione di* The Concept of a Person, in H. Putnam 1987, p. 155-156.

[4] E. Carli (2003), p. 82. Dennett ha affermato anche che «noi filosofi siamo più bravi a fare domande che a dare risposte. [Ma] escogitare gli interrogativi migliori, spezzando le vecchie abitudini e le tradizioni invalse nel modo di porre i problemi, è una parte molto difficile di quel grandioso progetto umano che è la comprensione di noi stessi e del nostro mondo» (D. C. Dennett 1997, p. 7).

[5] S. Nannini (2002), p. 113 (nota 27).

per anni»[6] dove in sostanza viene esaurito e risolto un grande enigma dell'umanità, e un'altra in cui il libro scritto sarà «affascinante, stupefacente, ma talmente difficile da comprendere, così ambiguo e oscuro che centinaia di filosofi, studenti, ricercatori e studiosi spenderanno le loro energie intellettuali nello sforzo di comprenderlo»[7] riscuotendo gloria e fama per centinaia d'anni. Per Dennett i filosofi sceglierebbero la seconda alternativa rifiutandosi di scrivere per queste ragioni un libro «fondamentale, chiaro e risolutivo»[8], al contrario degli scienziati che opterebbero, naturalmente, per la prima soluzione.

Addirittura secondo il funzionalista Philip Johnson-Laird «le teorie della mente dovrebbero essere espresse in una forma che possa essere tradotta in un programma di computer»[9], in modo tale da lasciare il minor spazio possibile all'«intuizione», rendendo accettabile l'idea per cui un modello filosofico, in questo caso della mente, sia valido solo se computabile da un calcolatore sotto forma di algoritmo. Questo ovviamente inasprisce i toni della disputa tra scienza e filosofia e Johnson-Laird radicalizza i problemi, complicandoli non lasciando traccia di vie risolutive. Efficace a questo proposito la battuta di Edoardo Boncinelli che sintetizza forse l'animo del nostro tempo, soprattutto in relazione ai problemi della filosofia della mente: per una spiegazione effettiva è «troppo presto per gli scienziati e troppo tardi per i filosofi»[10]. Boncinelli sottolinea dunque quella sensazione, talvolta condivisa da chi si occupa di scienza, della fine dell'indagine filosofica a favore del solo rigore dimostrativo della scienza, ma ciò non vuol dire che la scienza possa presumere di spiegare il reale, poiché di fatto non lo fa e probabilmente non può farlo, senza la filosofia. Di questo avviso forse non era Feyerabend quando affermava, in *Contro il metodo*, che nella scienza non c'è logica che la filosofia possa giustificare o criticare: «gli scienziati creano e sposano le teorie scientifiche per ragioni che sono in ultima analisi soggettive e perfino irrazionali»[11]. Ma ciò non pregiudica la rilevanza della filosofia della scienza o della filosofia *tout court*.

---

[6] E. Carli (2003), p. 83.

[7] Ibidem.

[8] Ibidem.

[9] P. N. Johnson-Laird (1988), p. 59. Riconosce però la difficoltà a livello algoritmico di formalizzare il pensiero. Afferma infatti, a proposito dei sistemi euristici, che «forse la maggior parte dell'attività di pensiero sta tra questi due estremi: ha un obbiettivo, ma non è svolta come un calcolo» (Ivi, p. 236).

[10] M. Di Francesco (2000), p. 11.

[11] J. Horgan (1998), p. 83.

Il rapporto tra scienza e filosofia è oggi rintracciabile nelle neuroscienze. Il confine tra le due attività umane per eccellenza sta proprio nell'approccio neurobiologico che caratterizza gran parte del panorama speculativo intorno alla mente. Il cosiddetto «problema mente-corpo», che ha lasciato tracce visibili in tutta la storia del pensiero, funge ora da spartiacque tra le indagini metafisiche e quelle prettamente scientifiche. Entrambe metodiche e raffinate, ma diverse nei loro obbiettivi. Come nota anche Jerry Fodor, «nell'ambito delle ricerche sulla mente, da alcuni anni si è diffuso un enorme interesse per lo studio del cervello: non solo i neurobiologi, ma anche molti filosofi e psicologi cognitivi, sembrano perdutamente innamorati di neuroni e connessioni sinaptiche, gangli basali, lobi prefrontali e cose del genere»[12]. I filosofi capiscono che la via per argomentare metafisicamente è costellata di nozioni scientificamente rilevanti, come ammette serenamente John Searle,[13] e al tempo stesso alcuni scienziati, talvolta addirittura premi Nobel come Gerald Edelman, non esitano a colorare di connotazioni filosofiche le loro teorie. Edelman stesso dice, d'accordo con Quine, che «la conoscenza comune, la conoscenza scientifica e l'analisi filosofica costituiscono un percorso continuo o almeno si intersecano in molti punti»[14]. Dunque questo accanimento contro le pretese filosofiche non è un motivo di fondo continuo nel campo della scienza. Anche se non condivide quelle teorie filosofiche che contrastano apertamente con la sua visione del reale, Edelman si dichiara aperto al rapporto con la filosofia, afferma infatti: l'interazione «non solo la ritengo possibile, ma credo sia essenziale, indispensabile. La scienza del cervello, e della mente, deve necessariamente stabilire delle relazioni con la filosofia, nel senso di una comune ricerca di armonie, di consonanze tra concezioni diverse. Non ritengo affatto la filosofia qualcosa di estraneo all'indagine scientifica, e la mia speran-

[12] E. Carli (2003), p. 142. Di fondamentale importanza è comprendere che il cervello «è fra gli oggetti più complicati dell'universo e certamente una delle strutture più straordinarie emerse con l'evoluzione. […] i suoi schemi di connessione, la sua dinamica, le sue modalità di funzionamento, e il suo rapporto con il corpo e con il mondo non ha eguali tra i vari oggetti o fenomeni in cui si è imbattuta la scienza» (G. M. Edelman-G. Tononi 2000, p. 45). Si può anzi dire senza timore di smentite che il cervello è «l'oggetto materiale più complicato dell'universo conosciuto» (G. M. Edelman 2004, p. 13).

[13] Cfr. E. Carli (2003), p. 189.

[14] Ivi, p. 116.

za è che i conflitti nelle diverse concezioni e ipotesi sulla mente possano risolversi»[15]. Putnam è probabilmente dello stesso avviso quando afferma che «ciò che conta è che, a dispetto degli stereotipi della scienza e della filosofia, ormai diventati paraocchi che impediscono la vista ai profani, scienziati e filosofi, scienza e filosofia sono attività interdipendenti: i filosofi hanno sempre trovato essenziale rifarsi alla conoscenza scientifica del loro tempo, e gli scienziati hanno sempre trovato essenziale fare una certa dose di filosofia nell'ambito del loro stesso lavoro scientifico, anche se rifiutavano di ammetterlo. Non mi sembra importante decidere se la scienza sia filosofia o la filosofia sia scienza, purché si abbia una concezione dell'una e dell'altra che le renda ambedue essenziali per una visione responsabile del mondo reale e del posto che l'uomo vi occupa»[16].

Rimanendo vero che «è il modo stesso di far scienza che può servire alla filosofia, che ha molto da imparare dal rigore, dall'onestà, dalla chiarezza nelle procedure e nei significati tipici del fare scienza»[17], è vero anche che essa, da sola, non può spiegare il reale. Ciò è confermato dal suo stesso evolversi, come analiticamente nota Horgan. «La scienza stessa, via via che progredisce, impone sempre nuovi limiti alle proprie possibilità: la teoria della relatività ristretta di Einstein vieta la trasmissione di materia e anche di informazione a velocità superiori a quella della luce; la meccanica quantistica stabilisce che la nostra conoscenza del microcosmo sarà sempre incerta; la teoria del caos conferma che molti fenomeni sarebbero impossibili da prevedere anche a prescindere dall'indeterminazione quantistica; il teorema di incompletezza di Gödel nega la possibilità di costruire una descrizione matematica completa e coerente della realtà; e la biologia evoluzionistica ci ricorda continuamente che siamo degli animali, plasmati dalla selezione naturale non per scoprire le profonde verità della natura ma per riprodurci»[18]. Come poi inevitabilmente accade, si sprofonda

[15] Ivi, p. 123. Afferma però: «la scienza, anche se riuscirà a reintegrare la mente nella natura, non sarà mai in grado di descrivere in modo adeguato l'esperienza individuale o storica. La scienza, tuttavia, dà una descrizione soddisfacente (di fatto, la migliore) dei *vincoli* che limitano l'esperienza» (G. M. Edelman 1993b, p. 252).

[16] H. Putnam (1987), p. 21. Come questa citazione rispecchia, le tensioni tra il campo filosofico e quello scientifico non sono presenti nella stessa misura in diversi paesi: in Italia per esempio sono più accentuate (vedi sopra l'affermazione di Boncinelli) rispetto ad altri paesi in cui appunto il rapporto tra filosofia e scienza è assai meno conflittuale (considerazioni suggerite dal Prof. G. Camardi).

[17] G. Boniolo-P. Vidali (2003), p. X.

[18] J. Horgan (1998), p. 20.

nell'assenza di senso, nel meccanicismo che inchioda, come nell'aforisma di Weinberg: «quanto più l'universo ci appare comprensibile, tanto più ci appare senza scopo»[19].

Nonostante Johnson-Laird cerchi di rassicurare i filosofi dicendo che «la scienza cognitiva non minaccia di ridurre i fenomeni sociali e culturali a spiegazioni nei suoi propri termini»[20], ammettendo la necessità di distinguere diversi livelli di spiegazione del reale, vale sempre la conclusione importante del biologo Steven Rose: «gli esseri umani, la loro mente e il loro cervello, non sono sistemi chiusi bensì aperti. Per conseguire la comprensione di noi stessi dobbiamo riconoscere quest'apertura e renderci conto che le scienze che possono spiegarne le conseguenze non sono più quelle della psicologia individuale e della neuroscienza, bensì quelle del collettivo di individui che compongono la società umana. Tali scienze – linguaggi – del collettivo sono ecologia ed etologia, sociologia ed economia, e non possono (nonostante le pretese imperalizzanti di alcuni sociobiologi e di alcuni neuroscienziati) venire ridotte semplicemente a scienze dell'individuo»[21].

E' vero anche che a volte è necessario dividere piuttosto che unire. E' il caso di Alberto Giovanni Biuso, che precisa come «*l'Artificial Intelligence* non è una questione da informatici, non è una tecnica fra le altre ma è il luogo in cui si esprime, oggi, l'invito inciso sul tempio di Apollo a Delfi: "conosci te stesso"»[22], inserendo non a torto il dibattito sull'intelligenza artificiale nel campo antropologico, con prospettive post-umane; aspetto questo assolutamente primario e fondante di ogni ricerca di una intelligenza non biologica: si tratta di metafisica, non di informatica. E' anche il caso di Diego Marconi, a cui interessa infatti precisare, giustamente, la differenza che intercorre tra filosofia e psicologia cognitiva. La psicologia cognitiva, vicina alle neuroscienze e ai loro risultati (le scienze cognitive), rappresenta infatti l'aspetto più «neuropsicologico» dell'indagine filosofica, al punto che Marconi la ritiene «indiscutibilmente una scienza empirica, e non un ramo della filosofia»[23]. Riconducendo però la scienza cognitiva a «un programma di ricerca [la cui] idea centrale è che ogni processo cognitivo può essere concepito come un processo di elaborazione di informazioni […] che può essere caratterizzata indipendentemente dalle proprietà

[19] Ivi, p. 120.

[20] P. N. Johnson-Laird (1988), p. 426.

[21] S. Rose (1994), p. 18.

[22] A. G. Biuso (2004), p. 73.

[23] D. Marconi (2001), p. 15.

materiali (fisico-chimiche, ma anche biologiche) del suo supporto»[24], si rivelano i limiti del cognitivismo stesso, proprio nel suo eludere l'apporto filosofico-critico necessario ad ogni scienza: manca il *senso*.

Ed è proprio alla ricerca del senso che Felice Cimatti dedica il suo prezioso volume *Il senso della mente*. Egli tenta di allargare il concetto di scienza in modo tale che essa possa comprendere la mente, conferendole partecipazione alla natura pur non essendo una «cosa». Così si esprime Cimatti: «una descrizione scientifica della mente umana non può essere né riduzionistica, perché in questo modo quel che si spiega non è la mente, né peraltro antinaturalistica, progetto che salvaguarda l'autonomia della mente solo al prezzo di trasformarla in una entità misteriosamente separata dal mondo naturale: la mente è naturale ma non è una cosa»[25]. E specificamente a proposito del senso aggiunge: «il senso non si misura né pesa, il senso non si riduce ad altro. Di più, non è il senso, l'intenzionale, che dev'essere spiegato mediante nozioni e strumenti non intenzionali, ma – al contrario – sono questi ultimi a presupporlo. L'insensato (ché questo significa, propriamente, non intenzionale) tanto non può rendere conto del sensato, quanto, in realtà, è intelligibile solo a partire da quest'ultimo. Non è il cervello che spiega la mente, ma – paradossalmente – è la mente che permette di pensare il cervello»[26]. Cimatti precisa ancora e in maniera profonda la questione. «Solo per una mente può esistere il non mentale, mentre il non mentale non può pensarsi in alcun modo, e quindi a maggior ragione non può pensarsi come non mentale, come mero fisico. Non solo, allora, non si può dare conto del senso, del contenuto, con il non sensato, con la pura e vuota forma; al contrario, può darsi qualcosa come il non sensato solo se presupponiamo il sensato. Non è la mente che deve essere spiegata con il fisico, ma è semmai il fisico che – per poter anche solo essere pensato come fisico – presuppone la mente, per quanto poi se ne dimentichi completamente, al punto di credere di essere il concetto fondamentale»[27]. Sembra proprio paradossale che il fisico "creda" di essere il concetto fondamentale mentre lo è il mentale, la mente intenzionale e dunque sensata; eppure Cimatti ne mette bene in mostra la realtà, da un punto di vista del senso. Il fisico non deve "dimenticare" il mentale, in quanto non solo la mente fa parte della natura ma addirittura la *significa*.

---

[24] Ivi, p. 18.

[25] F. Cimatti (2004), p. 14.

[26] Ivi, p. 77.

[27] Ivi, p. 110.

Nel caso della coscienza si pone un'ulteriore questione: se, come afferma Thomas Nagel, la ricerca scientifica si pone come esclusione del *soggettivo*, allora l'esperienza non avrà più «alcun luogo dove rifugiarsi»[28]. Se il limite della scienza secondo Cimatti risiede nel non comprendere che «il campionario delle entità naturali è assai più ampio del campionario di quelle materiali»[29] o causali, secondo Di Francesco sta nel fatto che «la natura stessa della metodologia scientifica la rende inadatta a trattare dei fenomeni coscienti»[30], ovvero pone l'alternativa importante tra dover pensare che «la scienza è incompleta perché 'lascia fuori' una dimensione della realtà, oppure che la nozione di esperienza cosciente è vuota e confusa perché non se ne può dare un'analisi scientifica»[31]. L'alternativa rappresenta un solco tra scienza e filosofia, che va oltrepassato per giungere a quell'interdipendenza di cui si diceva prima con Putnam e Edelman, ma al tempo stesso chiarendo la necessità sia di non "lasciar fuori" nulla dal campo del reale sia di conferire uno statuto ontologico al mentale.

John Searle ha espresso in modo chiaro, nel suo stile, la condizione della filosofia oggi, riassumendo i toni di questa introduzione. «Sono convinto che il senso autentico della filosofia stia ancora oggi nel porre problemi e nel tentativo di darne soluzione; molti dei problemi filosofici che preoccupavano Platone e Aristotele costituiscono tutt'oggi dei seri problemi filosofici. Dissento totalmente da coloro che sostengono che i nostri problemi provengono dal fatto che siamo giunti ormai alla fine della filosofia, alla morte della filosofia. Ho l'impressione invece che si sia appena iniziato a lavorare, a porre le questioni, e a porle nel modo corretto, e c'è sicuramente ancora un lunghissimo percorso da compiere, e credo sia proprio questo il significato più autentico del termine "filosofia". Continuare a porre, socraticamente, le questioni essenziali, fondative e ontologiche: che cos'è la mente, che cos'è l'intenzionalità, che cosa significa agire coscientemente, che cos'è la razionalità?»[32].

---

[28] M. Di Francesco (2000), p. 24.

[29] F. Cimatti (2004), p. 16.

[30] M. Di Francesco (2000), p. 24.

[31] Ibidem.

[32] E. Carli (2003), pp. 194-195.

«E' perché qualcosa degli oggetti esterni penetra
in noi che vediamo le forme e che pensiamo»
Epicuro[1]

Capitolo 1.

Il problema mente-corpo in filosofia

Volgendo lo sguardo ai filosofi e agli scienziati che hanno dedicato attenzione alla possibilità di una intelligenza artificiale, sia prima che dopo aver inteso questa come una vera e propria branca delle scienze umane, è possibile accorgersi della vastità e della complessità dei temi e delle teorie argomentate, per cui risulta inevitabile l'esclusione della trattazione di gran parte di queste dal presente lavoro. Pare tuttavia necessario cogliere i temi essenziali che hanno caratterizzato il maturare dell'attuale dibattito riguardante l'intelligenza artificiale, a partire proprio dagli albori della filosofia.

Il rapporto tra mente e cervello può essere ritrovato nell'antichità solo se visto nell'ottica di anima e corpo, dato che l'anima, come per noi la mente, era considerata quel «soggetto che conferisce unità agli atti e ai pensieri di una singola persona»[2], mentre il corpo era inteso come *organon*, cioè strumento dell'agire, non essendo noto neanche l'organo, come ora lo intendiamo, corrispondente al cervello.[3]

Come precisa Di Francesco, «all'uomo omerico sembrerebbero […] mancare molti caratteri che per noi sono essenziali della soggettività (non patologica) quali l'unità psichica, l'autonomia della volizione e la dimensione introspettiva»[4]; dunque lasciando da parte la concezione omerica per cui non esisteva neanche il concetto di anima come è stato definito sopra, la speculazione filosofica che frutterà la messe di teorie sul rapporto tra anima e corpo si dovrà essenzialmente agli atomisti e ai platonici. Dividiamo in due le correnti di pensiero solo per semplificare e non per sminuire o indebolire la complessità delle posizioni, che già agli albori del pensiero sono fittissime e di valore fondante. Gli atomisti e, con loro, gli stoici pensavano che la percezione fosse strettamente legata al pensiero considerando

---

[1] Epicuro, *Lettera a Erodoto*, citato da J. P. Changeux (1983), p. 151.

[2] S. Nannini (2002), p. 4.

[3] Cfr. J. P. Changeux (1983), p. 14, in cui si affronta la questione della collocazione dell'organo del sentimento e dell'intelligenza nei tempi antichi.

[4] M. Di Francesco (1996), p. 57.

l'anima come «il punto verso il quale confluiscono tutte le informazioni provenienti dai sensi e dal quale partono tutti i comandi che guidano il corpo nei suoi movimenti»[5]; essi riescono inoltre ad identificare nel cervello la base materiale del pensiero.[6] I platonici, convenendo con gli atomisti sulla tesi cefalocentrica, ritengono l'anima nettamente distinta dalla forma materiale del corpo attribuendo ad essa una natura immateriale. Si tratta dunque di *dualismo* ma non si escludono concezioni *pluraliste* in certi casi. Nella schiera di coloro che, a partire dal XVIII secolo, verranno chiamati "materialisti" compaiono Democrito, Epicuro, Lucrezio e in genere gli stoici, mentre tra i platonici si trovano, con le debite differenze, sia Agostino che, in qualche modo, Tommaso d'Aquino. Vi sono anche posizioni moniste, come l'idealismo di Plotino (l'anima non dipende dal corpo ma viceversa), e il caso aristotelico, che è complesso e riconducibile all'*ilomorfismo* (anima come forma del corpo), ma fortemente anti-platonico in questo senso. L'anima aristotelica è la *funzione* (la forma) di un corpo che ha la vita in potenza, ovvero è ciò che conferisce ad un corpo (e strettamente legato ad esso) che abbia gli organi adatti allo scopo (la vita in potenza, per la finalità naturale) la capacità di vivere, nel senso di sentire e muoversi.

È possibile ritrovare gli stessi approcci anche nel succedersi, nel tempo, delle teorie filosofiche. Di importanza fondamentale è il dualismo di Descartes, sulla scia dei platonici (in particolare Plotino in cui è espresso «per la prima volta in forma compiuta nella storia del pensiero filosofico il concetto di coscienza e d'interiorità come base del conoscere»[7] e Agostino cui Descartes deve quella «metafora della mente come 'spazio interno' accessibile solo a se stessi»[8]). La materia, per Descartes, non è più la «potenza» aristotelica capace di accogliere le forme, ma adesso è sostanza autonoma. Al dualismo va poi contrapposto il vero e proprio materialismo, di matrice atomista, di Hobbes. Egli critica Descartes e il suo *cogito*[9] e identifica la sensazione con l'evento mentale: è la teoria dell'identità mente-

[5] S. Nannini (2002), p. 13. Inoltre gli atomisti concepivano l'esistenza di due sole realtà: gli atomi e il vuoto. Non negando l'esistenza dell'anima, essi dunque affermano che anch'essa deve essere corporea. Invece gli stoici in realtà erano *ilozoisti* cioè ritenevano tutta la materia vivente colma di spirito (in ciò si distinguono dall'ateismo atomista).

[6] Cfr. J. P. Changeux (1983), p. 15.

[7] S. Nannini (2002), p. 15.

[8] Ivi, p. 16.

[9] Descartes aveva affermato che "per il fatto di stare pensando, io sono pensiero"; Hobbes lo critica affermando ironicamente che "non perché io sto passeggiando, devo essere una passeggiata".

cervello.[10]

Lo stesso dualismo si è presentato sotto varie forme tanto da poterne individuare cinque tipologie.[11] In primo luogo, esiste un dualismo *metodologico* per cui la differenza tra stati mentali e eventi naturali risiede nel loro modo di essere conosciuti. In secondo luogo, vi è un dualismo detto *concettuale intensionale* per cui a ogni predicato psicologico potrebbe corrispondere un predicato fisico con la stessa estensione. In terzo luogo, distinguiamo un dualismo *concettuale estensionale* che si dà come contrario dell'intensionale, ovvero «le descrizioni e spiegazioni psicologiche non sono traducibili in termini fisici o biologici»[12]. In quarto luogo, vi è il dualismo più noto ossia quello cartesiano: dualismo *delle sostanze*; esso prevede che pensiero e materia siano due sostanze completamente diverse e con proprietà distinte, tanto che dividono l'esistente in due parti, la *res cogitans* e la *res extensa*. In quinto e ultimo luogo, un dualismo diffuso, cioè quello detto *delle proprietà*, secondo cui pensiero ed estensione, pur restando diverse e irriducibili tra loro, sono caratteristiche «che possono appartenere simultaneamente alla medesima sostanza»[13].

Tutte queste forme si dividono e intrecciano nel modo di correlare il fisico al mentale: l'evento fisico può causare quello mentale e viceversa (*interazionismo* cartesiano), oppure non esiste relazione di causa ed effetto tra gli eventi ma interviene una causa esterna (*parallelismo* in cui rientrano le teorie di Spinoza, Malebranche e Leibniz, pur con differenze, e la teoria introspezionista di Wundt) oppure ancora l'evento mentale emerge da quello materiale ma il mentale non può avere efficacia causale nel mondo fisico (*epifenomenismo*).

La recente posizione popperiana ripresenta in vesti nuove il dualismo cartesiano delle sostanze. Popper definisce meglio la propria posizione dichiarandosi un pluralista (come si può definire anche la teoria platonica stessa), infatti propone la suddivisione del reale in tre «mondi» chiamati semplicemente *mondo 1* (corpi fisici con stati fisici), *mondo 2* (stati mentali o «disposizioni degli organismi» o «disposizioni al comportamento») e *mondo 3* («pro-

[10] Cfr. G. Gava (1983), p. 3.

[11] Se ne può contare, meno seriamente, una sesta per cui si considera *letterario* un dualismo che si presenta quando, in un romanzo, il personaggio creato dall'autore prende vita propria nel senso che l'autore non se ne può riconoscere artefice totale. E' un dualismo discusso in D. R. Hofstadter-D. C. Dennett (2001) a p. 374 e in altre parti, riferendosi ai casi di doppia personalità, e comunque in tono semiserio.

[12] S. Nannini (2002), p. 148.

[13] Ivi, p. 148. Un dualista di questo tipo è David Chalmers, di cui si parlerà nel terzo capitolo.

dotti della mente», appartenenti anche al mondo 1).[14] Popper ha collaborato col neurobiologo Eccles,[15] premio Nobel per la medicina, con cui ha scritto anche un testo – *L'io e il suo cervello* – dove si afferma che «il cervello e la mente esistono come entità separate, l'una con proprietà fisiche quali l'estensione e la localizzazione spaziale, l'altra immateriale, e interagiscono fra loro»[16] precisando che la mente cosciente «interagisce con alcuni moduli [«unità funzionali colonnari multicellulari» di qualche riga prima] del cervello, i cosiddetti "moduli aperti", i quali sono situati nell'emisfero dominante per il linguaggio, l'emisfero sinistro. La mente agirebbe sul cervello producendo delle lievi modificazioni dell'attività dei moduli aperti all'interazione, i quali occuperebbero le aree associative dell'emisfero sinistro»[17].

Si presentano a questo punto diverse posizioni non inquadrabili nel nostro schema essenziale: *l'empirismo* di Locke e il *monismo immaterialistico* di Berkeley, oltre la critica kantiana. Nel primo, il problema mente-corpo è ritenuto comunque insolubile,[18] mentre nel secondo, vicino alla posizione plotiniana e di Leibniz[19], viene difesa l'idea che esistano soltanto le sostanze spirituali (l'io che percepisce è la sola sostanza esistente).[20] Interessante la posizione di Kant che si preoccupa dell'unità della percezione e critica Descartes per aver con-

---

[14] Cfr. K. R. Popper (1996), pp. 15, 110, 148; cfr. inoltre G. Vallar (1981).

[15] Cfr. K. R. Popper (1996), p. 175 e ss. e cfr. G. Vallar (1981).

[16] Ivi, p. 151.

[17] Ibidem.

[18] Locke, Hume e Kant riterranno velati di mistero i problemi metafisici legati alla mente e al cervello. Locke ritenne che Dio stesso provvide, nell'atto della loro creazione, a garantire l'interazione mente-corpo e concepì la mente come "relazione di stati" appunto mentali (l'associazione di idee e di ricordi formano la "persona"). Hume propose invece la "teoria del fascio" (*bundle-theory*): la mente non è sostanza ma un «fascio di percezioni» interrelate che costituisce gli stati mentali. La mente è dunque gli stati mentali stessi. Il suo scetticismo riguardo la realtà "data" si unisce all'empirismo in quanto la coerenza e la costanza dell'esperienza sensoriale costringono il filosofo a ritenere reale la percezione. Kant affermò invece che materia e spirito devono essere sostanze di tipo diverso, almeno dal punto di vista fenomenico (cioè l'orizzonte invalicabile dell'esperienza).

[19] Leibniz infatti concepiva alla base della realtà le *monadi*, atomi "spirituali". Così differenzia l'anima dal corpo postulando che la prima è composta di monadi coscienti mentre il secondo di monadi deboli e inconsce. Il famoso esperimento mentale del "mulino" leibniziano confermerebbe la non riducibilità dell'anima alla materia. Cfr. S. Nannini (2002), pp. 28-29.

[20] Hegel si può accostare al monismo immaterialistico con la seguente precisazione: per lui, la materia va negata solo da un punto di vista filosofico, in quanto l'anima è il punto in cui nel corpo l'«Idea trova se stessa come spirito», ma non da un punto vista scientifico (che comunque ritiene subalterno a quello filosofico).

fuso l'autocoscienza con la conoscenza di una sostanza indipendente dall'esperienza, mentre per Kant «non esistono esperienze senza soggetto»[21], in quanto il soggetto è ricettacolo di esperienze non unitarie ma che necessitano della presenza di un'unità di coscienza oltre che di un rapporto tra io empirico e mondo esterno (entità necessariamente differenziate): questo è l'*idealismo trascendentale*. L'anima, insieme al mondo e a Dio, è un concetto metafisico inintellegibile, dato che l'intelletto cadrebbe in «paralogismi» ogni qualvolta tentasse di coglierne la natura. Per questo motivo, identificando il soggetto con il corpo, Kant ritiene che «diviene io il luogo là dove "si sente", cioè il "si" del sentire diviene "io" in questo sentirsi-sentire, e in questo conoscer-si e riconoscer-si (e quindi in questo venire a coincidere con sé) come quello stesso "si" che sente e che "si" sente»[22].
La reale portata dell'intervento kantiano si può percepire nel momento in cui si coglie l'attributo fondante del suo *Io penso*: non è una sostanza come per Descartes né tantomeno una "collezione di percezioni" come per Hume, bensì è una *funzione*, concetto fondamentale che verrà ripreso più avanti, «*funzione di unificazione*, di sintesi, di tali percezioni»[23].

Con la teoria dell'evoluzione biologica e segnando così la fine del periodo in cui si afferma la concezione del cervello come «organo» del pensiero, ma al tempo stesso indicando l'inizio, lo sviluppo «embrionale», dell'affermarsi della scienza nelle teorie della mente, Darwin rivoluziona gli assetti teorici superando il *fissismo* di matrice aristotelica. A partire da questo fervente periodo di rigenerazione culturale e certamente a seguito di notevoli diramazioni contestuali nonché interdisciplinari di certe idee, si è assistito al proliferare delle posizioni da cui emergeranno essenzialmente la *teoria dell'intenzionalità*, il *funzionalismo psicologico* e il *monismo neutrale*. La prima risale a Brentano secondo cui gli stati mentali hanno, come particolarità necessaria, l'intenzionalità ovvero l'essere diretti verso qualcosa; caratteristica, questa, assente negli stati fisici che così vengono dualisticamente distinti da quelli mentali. La seconda posizione si concretizza nel *pragmatismo* di Peirce, James e poi Dewey in base al quale la conoscenza è «una funzione biologica di adattamento all'ambiente»[24] e, identificando mente e coscienza come in Descartes, gli stati mentali sono

[21] M. Di Francesco (1996), p. 71.

[22] F. Leoni (20029, p. 70.

[23] S. Nannini (2002), p. 44.

[24] Ivi, p. 63. Riprenderemo brevemente l'argomento alla fine del secondo capitolo.

«stati interni accessibili introspettivamente»[25], pur non concedendo all'introspezione particolare rilevanza.[26] La terza, sostenuta in precedenza anche da Hume e ripresa poi successivamente sia dallo stesso James che da Russell, fa riferimento a Mach che riteneva l'esperienza l'unico orizzonte della realtà, costituita da elementi fisici o mentali a seconda di come vengono interpretati. Anche altri pensatori hanno accostato le loro posizioni filosofiche al monismo neutrale, chi subito chi invece solo come approdo finale dopo un diverso itinerario teorico.[27]

In seguito Freud svelerà l'aspetto inconscio degli stati mentali senza distanziarsi particolarmente dalla posizione di James: il rapporto mente-corpo è una delle questioni che Freud dibatte senza sosta, rimeditando e riformulando la propria concezione per l'intero corso della vita. La continuità della vita mentale è un punto di partenza per Freud, che nelle prime opere, della fine del XIX secolo, presenta ancora in maniera incerta: egli «chiama questa vita mentale *psiche*, perché ancora incerto sui rapporti che legano mente e corpo, così da pensarli inizialmente in un'ottica di *salto* (di "misterioso salto tra la mente e il corpo") e, poi, di interazione […]. L'idea di una continuità nella vita mentale di ciascuno – l'idea che passato e presente, giorno e notte, vita e morte siano tra loro ben più connessi di quanto la coscienza possa fare credere; che le primissime fasi della crescita si leghino strettamente a quelle successive – è di essenziale importanza per comprendere lo sviluppo della mente e le sue "deviazioni", come pure per apprezzarne i dinamismi, i conflitti, i processi»[28]. Inizialmente si parla dunque di *connessioni* tra trauma e sintomo, tra psichico e somatico; il bilanciamento nell'atteggiamento di Freud tra biologia e psicologia permane costante, «sembra comunque evidente che per il fondatore della psicanalisi, così come per tanti altri studiosi dopo di lui, la mente *poggia* sul corpo ed è questo corpo che orienta la mente [difatti] l'Io è innanzitutto un Io corporeo, dirà Freud nel 1922»[29].

A questo punto la ricerca scientifica si afferma nettamente sulla speculazione filosofica, a volte prendendone il sopravvento. Soprattutto negli ultimi anni si è generata un'«espansione a macchia d'olio» delle posizioni filosofiche, ma inizialmente contraddistinguono il XX se-

[25] Ivi, p. 64.

[26] Si parlerà nel secondo capitolo del pragmatismo, nelle vesti del funzionalismo psicologico.

[27] Di questo approccio si tratterà diffusamente nell'ultimo capitolo.

[28] A. Lis-S. Stella-G. C. Zavattini (1999), p. 16.

[29] Ivi, p. 20.

colo due correnti filosofiche fondamentali: la *fenomenologica* e l'*analitica*. L'approccio fenomenologico di Husserl riprende per grandi linee l'idealismo trascendentale di Kant e per certi versi il monismo neutrale già incontrati. «Il soggetto è l'occhio che vede le cose, ma che proprio perciò non può vedere se stesso come una cosa tra le altre. Pertanto il soggetto non fa parte del mondo reale. […] Una persona, in quanto corpo concepito come 'macchina' puramente fisica (ossia in quanto *Körper*), esemplifica l'essenza della Natura Materiale; in quanto corpo vivente proprio (ossia in quanto *Leib*) esemplifica invece l'essenza della Natura Animata. […] Essa è dunque un'unica sostanza che esemplifica molteplici essenze»[30]. Secondo Arendt «la fondamentale e più grande scoperta di Husserl coglie con la massima precisione l'intenzionalità di ogni atto della coscienza, il fatto, cioè, che nessun atto soggettivo è mai senza un oggetto […]. In virtù dell'intenzionalità l'oggettività è incorporata nella soggettività stessa della coscienza»[31]. L'approccio analitico invece, soprattutto per quanto concerne la sua parte iniziale definita "classica", si concentra sull'analisi linguistica. Certamente il fenomeno analitico è poliedrico e presenta facce che poco o nulla possono riguardare la filosofia della mente o il rapporto mente-corpo. In realtà è però possibile ricondurre certe posizioni analitiche a ben precise teorie del mentale, nonostante a volte certe «riconduzioni» si dimostrino delle «riduzioni». Dirigendo perciò lo sguardo al mentale, l'analisi linguistica iniziata con Frege approderà al *comportamentismo logico* ad opera soprattutto di Carnap ed Hempel, con atteggiamenti di stampo materialista, nonché *fisicalista* (riducibilità di tutte le scienze alla fisica). Wittgenstein e Ryle si schierano contro il dualismo (per usare termini cari alla filosofia della mente) ma al tempo stesso si distanziano anche dal comportamentismo logico per approdare al *comportamentismo analitico*: «non c'è bisogno di degradare l'uomo a macchina per negare che sia un corpo 'abitato' da una mente; basta pensare che è l'uomo in carne e ossa a esser capace di un comportamento intelligente»[32]. Con la riflessione di Ryle prendono forma sia una concezione del rapporto mente-corpo totalmente velata di mistero (per cui la soluzione di esso è addirittura inintelligibile)[33] sia un *solipsismo* in cui il dubbio sull'esistenza delle altre menti, al di fuori della propria, diviene legittimo. Bechtel a questo proposito chiarisce i termini. Il comportamentismo psi-

[30] S. Nannini (2002), pp. 67-68.

[31] H. Arendt (1987), p. 128.

[32] S. Nannini (2002), p. 89, che riprende Ryle.

[33] Di ciò si parlerà meglio nel prossimo capitolo.

cologico è «un programma di ricerca empirica che cerca di scoprire le leggi per spiegare il comportamento degli esseri umani e degli altri organismi in termini degli stimoli cui sono sottoposti e della storia del condizionamento passato dell'organismo. La sua caratteristica consiste nel rifiutarsi di ricorrere agli eventi mentali per spiegare il comportamento»[34]. Invece il comportamentismo filosofico, frutto del lavoro di Ryle e del primo Wittgenstein incrociato al positivismo logico, «si occupa principalmente della semantica del nostro usuale vocabolario mentalistico [ovvero] di tradurre i termini che pretendono di riferirsi all'attività mentale in termini che menzionano soltanto comportamenti o disposizioni a comportarsi in certi modi»[35]. Infatti «Wittgenstein ritiene che molte questioni filosofiche, come quella relativa ai rapporti mente-corpo, siano il risultato di una confusione linguistica»[36]. Ma, riguardo al solipsismo, va detto che «tutte le teorie solipsistiche si trovano a discordare così intimamente dai dati della nostra esistenza e della nostra esperienza – sia che pretendano radicalmente che solo l'io "esiste" sia che, più moderatamente, sostengano che l'io e la sua coscienza di sé sono gli oggetti primari di ogni conoscenza verificabile»[37] – a causa del mancato riconoscimento tacito da parte degli altri, senza il quale appunto «non potremmo nemmeno prestar fede al modo in cui appariamo a noi stessi»[38]. L'apporto essenziale fornito dai filosofi e gli scienziati del Circolo di Vienna è stato l'avere ristretto l'indagine filosofica e scientifica dal «conoscibile in sé» a «ciò che si può conoscere» in maniera empirica. Solo in questa prospettiva è stato possibile che il cognitivismo e l'Intelligenza Artificiale nascessero, come progetti che indagano la mente da un punto di vista empirico.

Nel prossimo capitolo verranno esposte le tappe principali del cammino percorso dall'Intelligenza Artificiale, tenendo in considerazione i suoi rapporti con le teorie della mente più importanti in modo da completare l'essenziale quadro teorico fin qui esposto.

---

[34] W. Bechtel (1992), p. 146.

[35] Ibidem.

[36] Ivi, p. 147. Infatti Wittgenstein, con l'analisi del linguaggio ordinario, classifica i problemi della filosofia della mente come «crampi mentali» o «fraintendimenti linguistici».

[37] H. Arendt (1987), p. 129.

[38] Ibidem. Cfr. D. C. Dennett (1997), pp. 11-29.

«E le vecchie e tenaci credenze negli spiriti, nelle anime e nelle essenze, che cosa sono? Sono tutte insinuazioni che siamo impotenti a migliorare noi stessi. Cercare le nostre virtù in siffatti pensieri è altrettanto sbagliato, mi sembra, che cercare l'arte nelle tele grattando via l'opera del pittore»

M. Minsky[1]

«Pur non essendo né stupido né ignorante, egli agiva come se lo fosse»

A. R. Damasio[2]

Capitolo 2.

Storia dell'Intelligenza Artificiale e teorie della mente

Si può dire che «tutto ebbe inizio con i tentativi di meccanizzare i processi mentali di ragionamento. Si è spesso sostenuto che la nostra capacità di ragionare è ciò che ci distingue dalle altre specie: sembra quindi a prima vista un po' paradossale meccanizzare ciò che è più specificamente umano. Eppure perfino gli antichi greci sapevano che il ragionamento è un processo strutturato e che, almeno parzialmente, è governato da leggi esplicitabili. Aristotele codificò i sillogismi, Euclide codificò [(assiomatizzò)] la geometria; ma dopo di loro dovettero passare molti secoli prima che ci fosse un nuovo progresso nello studio del ragionamento assiomatico»[3]. E proprio nella ricerca dell'assiomaticità del ragionamento si incroceranno le strade della logica e della matematica costituendosi nella *metamatematica* o *metalogica*, quando il teorema di Gödel demolirà il programma di Hilbert. Questi riteneva che «fosse possibile trovare una dimostrazione della coerenza o della completezza che dipendesse soltanto da metodi «finitistici» di ragionamento, ossia da un numero ristretto di metodi comunemente accettati da tutti i matematici»[4]. Il primo teorema di Gödel, di incompletezza per sistemi corretti, recita: «*sotto certe ipotesi,un sistema corretto è incompleto* (sia semanticamente sia sintatticamente)»[5]; mentre il secondo recita: «*sotto certe ipotesi, un si-*

[1] M. Minsky (1989), p. 71.

[2] A. R. Damasio (1995), p. 76.

[3] D. R. Hofstadter (1984), p. 20.

[4] Ivi, p. 25.

[5] P. Odifreddi (2003), p. 222.

stema corretto non dimostra la propria correttezza»[6]. «In breve, Gödel metteva in evidenza che la dimostrabilità è una nozione più debole della verità, indipendentemente dal sistema assiomatico considerato»[7]. Aggiunge inoltre Williams che «se si fosse provato che la matematica era sia *coerente* sia *completa* – ovvero nessuno dei suoi assiomi fondamentali ne contraddiceva un altro o lasciava spazio per scappatoie – ci si sarebbe trovati vicino a provare la visione platonica e kantiana della matematica come collegamento innato dell'uomo con l'infinito»[8].

La ricerca dell'intelligenza artificiale è strettamente intrecciata con la logica. E' infatti a partire da quest'ultima che sorge la domanda sulla mente formulata nei termini di «leggi del pensiero»; nel cercare di capire «cosa *fa* la mente quando dimostra»[9] la logica dà vita all'Intelligenza Artificiale. Naturalmente quello di «intelligenza artificiale» è un concetto moderno ma sin dall'antichità vi era grande interesse nel concepire, e in certi casi realizzare, un prototipo meccanico che fosse in grado di *rappresentare* la conoscenza. L'*Ars combinatoria* fu probabilmente il primo tentativo in questa direzione, mettendo a frutto i principi dell'*Ars magna* di Raimondo Lullo, consistente nel «disporre su un circolo gli elementi fondamentali che formano una nozione […]. Queste caselle concettuali vanno poi messe in relazione fra loro da particolari schemi grafici che si possono disegnare al centro delle ruote»[10]. Essendo così ogni ragionamento un rapporto tra nozioni, si *geometrizza* la rappresentazione della conoscenza. Lullo costruì alcune *macchine inferenziali* di questo genere con l'intento di dimostrare la verità o falsità di un'asserzione, nel suo caso per scopi teologici e apostolici. Fu invece per districare problemi di rotte marinare che, sviluppando un'idea di Nepero, Gunter realizzò la *Tavola di logaritmi meccanica* (di cui si ha notizia non prima del 1624).[11]

Ma, come fa notare Pratt, l'idea di intelligenza artificiale va considerata come del tutto moderna, essendo necessariamente intrecciata alle condizioni storiche e culturali di quel perio-

---

[6] Ivi, p. 223. Si intende «coerenza: quando ogni teorema, se interpretato, risulta vero (in qualche mondo immaginabile)[; e] completezza: quando tutti gli enunciati che sono veri (in qualche mondo immaginabile), e che possono essere espressi come stringhe ben formate del sistema, sono teoremi» (D. R. Hofstadter 1984, p. 110).

[7] Ivi, p. 19.

[8] S. Williams (2003), p. 27.

[9] M. Di Francesco (1996), p. 132.

[10] N. Ubaldo (2003), p. 258. E cfr. V. Pratt (1990), p. 93.

[11] Ivi, p. 54.

do in cui la matematica, la logica e l'ingegneria presero connotazioni precise e strade molto vicine tra loro, cioè l'età moderna. L'intuizione che renderà possibile concepire una «macchina pensante» consiste proprio nella rappresentabilità della conoscenza, nel considerare la conoscenza come una rete di proposizioni, concezione fatta propria dalla logica proposizionale e predicativa. «Dobbiamo però – dice Pratt – opporci all'idea che, fin dall'inizio di questo periodo, esistesse un *piano*, un'ambizione comune che ispirasse tutti coloro che, in un modo o nell'altro, portarono il calcolatore moderno e il "pensiero artificiale" che esso sostiene più vicino alla possibile realizzazione. In questo senso non c'è una "storia" dell'avvento dell'intelligenza artificiale. Notiamo invece la presenza di una quantità di progetti diversi, alcuni di successo ed altri falliti, la cui importanza è percepibile dalla nostra prospettiva attuale in quanto noi ora possiamo giudicare cosa sarebbe dovuto accadere, quali fondamenti concettuali dovessero essere posti, prima che si potesse progettare, costruire e programmare il *calcolatore intelligente*»[12].

Leibniz, di cui si è già accennato prima per l'impegno filosofico, impiegò il suo genio nel tentare l'accostamento di ragionamento non matematico a quello matematico, con lo scopo di rendere plausibile la costruzione vera e propria di una macchina che *ragiona* logicamente, sfruttando l'uso che della matematica si era cominciato a fare nel periodo in cui egli visse. Il calcolatore di Leibniz[13] non fu sviluppato per centocinquant'anni ovvero fino a quando si rinnovò l'interesse per la matematica in quanto struttura, con Babbage intorno il 1830. Questi ideò una «macchina alle differenze», poi realizzata nella *macchina analitica*, con la caratteristica innovativa di «usare i propri dati in uscita come ingresso per un calcolo successivo»[14]. Ma la macchina di Babbage non fu costruita se non, in maniera ridotta, dal figlio.

Tappa fondamentale nel progresso dell'ingegneria delle macchine fu senza dubbio l'uso dell'elettricità che, a partire dal 1881 con Marquand (migliorando il progetto di Jevons) e

[12] Ivi, pp. 11-12.

[13] Leibniz trovò un difetto negli orologi calcolatori di Schickard e Pascal: la moltiplicazione era possibile solo mediante addizioni ripetute; inventò così il tamburo differenziato che fu usato nella prima macchina calcolatrice (di Thomas).

[14] Ivi, p. 141. La macchina analitica di Babbage era costituita da un «processore centrale» o «macina» e un «magazzino» o memoria in cui «i numeri sono rappresentati, come nella "macchina alle differenze", su colonne di ruote, soltanto che ora queste sono costituite da 41 elementi, e la ruota più in alto rappresenta il segno del numero memorizzato in quelle sottostanti» (Ivi, pp. 142-143).

poi con Aiken (col suo progetto *Mark I*) intorno al 1930, fu impiegata inizialmente come sostituzione di collegamenti meccanici ma, in seguito, sfruttando la tesi di Boole «in base alla quale il calcolo proposizionale era isomorfo[15] al calcolo dell'aritmetica binaria»[16], si fece uso del relé, anche se ancora con esso non si poteva realizzare alcun processo aritmetico bensì si poteva *trasferire numeri*. Successivamente «si trasferì l'uso del relé dal sistema telefonico al calcolatore automatico – non per collegare elementi meccanici, come nel Mark I di Harvard, ma al loro posto. Anziché un contatore a ruota, che rappresentava una conta di dieci mediante una rivoluzione completa, ed i numeri da uno a nove per mezzo di una rotazione parziale, il calcolatore della Bell utilizzava un relé, capace di assumere solo due stati (essenzialmente: aperto o chiuso)»[17]. Si tratta dell'*ENIAC* («il primo grande e veloce calcolatore, *general purpose* e programmabile»[18] di Eckert e Mauchly, che faceva uso del *flip-flop*, un particolare tipo di relé), cui seguirono l'*EDSAC* (di Wilkes, Renwick, Wheeler e altri) che presentava un programma memorizzato, il *MADM* (di Williams, Kilburn, Newman e in parte Turing), l'*ACE* (di Alway, Davies, Wilkinson e Woodger su progetto di Turing del 1945), l'*EDVAC* (di Snyder, Gluck e Boghosian su progetto di von Neumann[19]) e infine l'*UNIVAC* (degli autori dell'ENIAC e del BINAC) che fu «il primo grande calcolatore elettronico prodotto per scopi commerciali»[20]. I calcolatori moderni erano caratterizzati da un

[15] L'isomorfismo viene definito come «una trasformazione che conserva l'informazione» in D. R. Hofstadter (1984), p. 9.

[16] V. Pratt (1990), p. 165. Citando M. Di Francesco (1996) a p. 133, si può aggiungere che per Boole «processi mentali e algebrici hanno qualcosa in comune; ma che cosa e perché? Boole non offre alcuna risposta. Alla sua intuizione non corrisponde una teoria soddisfacente. Forse, egli suggerisce, trovarne una è un compito troppo complesso per le nostre facoltà intellettuali».

[17] V. Pratt (1990), p. 189.

[18] E. Burattini-R. Cordeschi (2001), p. 19. Per Williams, «il primo computer moderno» (S. Williams 2003, p. 17). Cfr. anche B J. Copeland, *Computation*, in L. Floridi (2004), pp. 3-4.

[19] Per von Neumann «tutta la vita può essere ridotta a sistemi autoriproduttivi che seguono regole matematiche fisse» (S. Williams 2003, p. 18). Un suo testo del 1948 «applicava al regno della biologia il concetto della macchina di Turing [di cui parleremo tra poco], presentando il genoma [anche se von Neumann non avrebbe potuto utilizzare questo vocabolo] umano come un sistema di calcolo programmato per riprodurre copie di sé stesso. Scritto diversi anni prima dell'analisi di Watson e Crick delle molecole del DNA, il testo di von Neumann anticipava il meccanismo chimico di trascrizione del DNA. Costituì anche il punto di partenza di una nuova scienza: la vita artificiale» (Ivi, p. 38).

[20] E. Burattini-R. Cordeschi (2001), p. 19.

programma memorizzato che permetteva l'autostrutturazione necessaria per l'I.A., grandi memorie ad accesso elettronico grazie al passaggio da dispositivi elettromeccanici ad elettrici, l'uso dell'aritmetica binaria basando le operazioni non su contatori bensì su elementi della logica booleana e, infine, l'esecuzione in modo centralizzato delle operazioni fondamentali. La carta fu sostituita dal nastro magnetico e il triodo dal transistor.

L'apporto di Alan Turing all'impresa dell'intelligenza artificiale va considerato come fondamentale. Come Pratt chiaramente illustra, è seguendo le linee di sviluppo della psicologia degli inizi del Novecento, intessuta dello sperimentalismo di Watson che in *Behavior*[21] «raccomandava alla psicologia di restringere la sua attenzione esclusivamente al fenomeno del comportamento»[22] cui fa seguito il comportamentismo di Skinner del 1938, che nel 1950 Turing «propose di interpretare la domanda "Le macchine possono pensare?" senza far riferimento ad alcun loro processo "mentale"»[23]. E' ancora Pratt a chiarire il ruolo svolto da Turing in questo ambito: «Il suo lavoro in logica lo aveva condotto a sviluppare una chiara descrizione di cosa fosse una procedura meccanica, e lo aveva condotto alla formulazione del corollario per il quale tutte le macchine a stati finiti (perché in grado di compiere un repertorio fondamentale di azioni semplici) erano logicamente equivalenti. Data la concezione predominante dell'organismo complesso con una gamma di comportamenti appropriati, fu possibile concepire il cervello come una macchina a stati finiti con funzione di controllo del comportamento. Non parve quindi assurdo credere nella possibilità di costruire un giorno un equivalente meccanico o elettronico»[24]. La macchina potrebbe essere dunque programmata allo scopo di «imitare il funzionamento cerebrale»[25]: si deve infatti a Turing la concezione di un calcolatore «idealizzato»[26], la nota *Turing Machine* che, basandosi sull'elaborazione di un dato in ingresso considerato il suo stato interno e scandendo ogni e-

[21] Pratt si riferisce a: J. B. Watson, *Behavior: An introduction to Comparative Psychology*, New York, Holt, Rinehart, Winston, 1914 (cfr. V. Pratt 1990, p. 232).

[22] Cfr. Ivi, p. 222.

[23] Ibidem.

[24] Ivi, p. 232.

[25] Ivi, p. 235. Pratt aggiunge inoltre riguardo Turing, a p. 302, che «benché il cervello possa apparire estremamente sofisticato, nel momento in cui lo si consideri una macchina a stati finiti, si può ritenere fattibile il progetto di costruirne un equivalente elettronico».

[26] Interessante notare, come fa Williams (cfr. (S. Williams 2003, p. 29), che Turing in inglese usa il termine «computer» essenzialmente come in italiano diremmo «ragioniere».

laborazione in stati finiti, sarebbe in grado, se dotata di memoria e tempi di calcolo indefiniti, di computare qualsiasi «funzione calcolabile mediante un procedimento algoritmico», secondo la tesi di Church-Turing.

In Burattini-Cordeschi si legge: «Gli aspetti fondamentali di questa descrizione informale e intuitivamente visualizzabile di una macchina di Turing possono essere espressi in un linguaggio più astratto e rigoroso. Si consideri la lista di simboli s_1, s_2, s_3, …; q_1, q_2, q_3, …; *D*, *S*. *Un'espressione* è una successione finita di simboli presi da questa lista. Un'*istruzione* è una *quintupla*, cioè un'espressione della forma $q_i s_j s_k q_l$M dove M sta per la lettera *D* o per la lettera *S*. Intuitivamente, un'istruzione si interpreta in questo modo: se la configurazione interna è q_i e il simbolo osservato è s_j, allora la macchina scriverà al suo posto il simbolo s_k entrando nella configurazione interna q_l e spostandosi di una casella, a destra o a sinistra, a seconda che M sia *D* oppure *S*. Si conviene generalmente che il simbolo s_1 indichi la casella vuota; pertanto rimpiazzare s_j con s_1 equivale all'azione di cancellare s_j. *Definizione*: Una macchina di Turing è un insieme finito (ma non vuoto) di quintuple che non contiene nessuna coppia di quintuple che coincidano nei loro primi due simboli»[27]. Accolta la definizione formale della MT, ci interessiamo ora alla descrizione più puntuale delle sue parti, avvalendoci dell'aiuto di Marconi che ne descrive compiutamente la struttura.[28] Una MT è costituita da un nastro di lunghezza idealmente infinita diviso in caselle e di un «cursore» che si sposta di casella in casella con capacità di lettura o scrittura, per ciascuna di esse, di un simbolo per volta. Il cursore può trovarsi in un numero finito di *stati interni*, compreso uno *stato finale*. La macchina esegue istruzioni con la forma di quintuple (insiemi ordinati di cinque elementi) di cui abbiamo già detto sopra. Marconi chiarisce sinteticamente che «in ciascuno stato in cui può trovarsi e per ciascun simbolo che può accaderle di leggere, la macchina 'sa cosa fare', e c'è una sola cosa che può fare»[29]. Naturalmente nessuna istruzione può cominciare con lo "stato finale" di cui sopra. Utile ora ricorrere alle precisazioni fondamentali che Tabossi fa nel suo testo.[30] La macchina può compiere solo tre operazioni, ovvero: *1.* leggere il simbolo della casella in esame; *2.* scrivere un altro simbolo al posto di quello; *3.* spostare il nastro di una casella a destra o a sinistra. Per eseguire qualsiasi istru-

[27] E. Burattini-R. Cordeschi (2001), p. 20.

[28] Per la descrizione della MT che segue, cfr. D. Marconi (2001), pp. 38-43.

[29] Ivi, p. 38.

[30] Cfr. P. Tabossi (1988), p. 42. Cfr. anche B. J. Copeland, *Computation*, in L. Floridi (2004), pp. 4-6.

zione nella macchina di Turing, è necessario che tali istruzioni specifichino: *a.* lo stato in cui si trova la macchina, esprimibile convenzionalmente con un numero (ad esempio, una certa configurazione d'apertura o chiusura degli interruttori di un circuito elettrico); *b.* il simbolo nella casella in esame (1 o 0); *c.* l'operazione da eseguire (1, 2 o 3 cui sopra); *d.* lo stato in cui si troverà la macchina a operazione avvenuta. Marconi[31] menziona un aspetto importante della MT e della dicitura di "stato interno" che può indurre in errore concettuale, cioè: la dizione 'stati interni' può far pensare che 'dentro' la macchina succeda qualcosa, che essa abbia una struttura interna cha và al di là del nastro e del cursore con le sue capacità di lettura, scrittura e spostamento, ma non è così dato che gli stati interni servono soltanto a far dipendere l'azione della macchina oltre che dal simbolo che legge anche dal "momento" in cui lo legge ovvero dalla fase della computazione in cui lo legge. Aggiunge che «se non avessimo a disposizione gli stati, alla lettura di uno stesso simbolo corrisponderebbe sempre la stessa azione della macchina; e la macchina non potrebbe terminare nessuna computazione, ma continuerebbe inesorabilmente a scrivere (o eventualmente cancellare) simboli, a seconda del simbolo che legge»[32]. Ritengo utile aver fatto così specifico riferimento, anche con dettagli, alla macchina di Turing, in quanto essa è da considerarsi ponte di contatto tra la scienza dell'informazione e le indagine filosofiche. La rilevanza di Turing non è certamente di stampo filosofico, ma il suo ruolo nel campo della scienza va interpretato con spiccati toni filosofici, specie comportamentisti.

Infatti «è possibile dimostrare che le funzioni calcolabili da una MT [Macchina di Turing] sono equivalenti a quelle comprese in una importante ed estesa classe di funzioni matematiche, dette funzioni ricorsive, le quali, a loro volta, secondo la cosiddetta "tesi di Church" sono equivalenti alle funzioni intuitivamente computabili. Sebbene la tesi di Church sia una congettura non dimostrabile formalmente (dato che cerca di rendere conto dell'idea intuitiva di "calcolabile in modo meccanico"), il credito di cui gode autorizza a concludere che una MT è in grado di calcolare ogni funzione computabile»[33]. Anche Frixione (1994) nota come le espressioni «algoritmo» e «funzione computabile in modo algoritmico» in realtà siano «concetti intuitivi, non specificati i modo formale, per cui una dimostrazione rigorosa di equivalenza con il concetto di funzione calcolabile da una MT non è possibile. La Tesi di

[31] Cfr. D. Marconi (2001), pp. 40-41.

[32] Ivi, p. 41.

[33] M. Di Francesco (1996), p. 140.

Church quindi non è un teorema, e neppure una congettura che potrebbe un giorno, in linea di principio, diventarlo. Si tratta di un enunciato che si fonda su un ampio spettro di evidenza matematica di tipo euristico»[34]. Continua Frixione affermando che alcuni studiosi considerano la Tesi di Church, non essendo mai stata confutata, «come una sorta di "legge empirica" piuttosto che come un enunciato a carattere logico-formale. Il logico Emil Post, il quale propose un concetto di macchina calcolatrice in parte analogo a quello sviluppato da Turing […], sottolineava il suo disaccordo con chi tendeva ad identificare la Tesi di Church con un assioma o una mera definizione [giacché andava considerata] una "ipotesi di lavoro", che, se opportunamente "corroborata", dovrebbe assumere il ruolo di una "legge naturale", una "fondamentale scoperta circa le limitazioni del potere matematizzante dell'*Homo sapiens*"»[35].

Le macchine che riescano a simulare ogni tipo di macchina di Turing sono dette «Macchine di Turing *universali*». Ma non tutto può essere computato, come ricorda Johnson-Laird, «poiché esistono molti problemi che possono essere enunciati ma non hanno alcuna soluzione computabile. E' impossibile ad esempio progettare una macchina universale in grado di determinare se una qualsiasi macchina di Turing selezionata arbitrariamente, sulla base di alcuni dati selezionati arbitrariamente, si fermerà o continuerà a computare per l'eternità. Non c'è alcun test che permetta di decidere in maniera sicura se un problema ha o no una soluzione computabile»[36]. L'interpretazione comportamentista di Pratt del *gioco dell'imitazione* di Turing, che abbiamo riportato prima, cioè la proposta che le macchine pensino anche solo se riescono a farlo credere a chi facesse loro delle domande e ricevesse risposte per via elettronica, è senza dubbio condivisibile e infatti, come aggiunge Frixione (1994), «ci troviamo di fronte ad una sorta di versione "estremista", o "radicale", della Tesi di Church, che, grosso modo, potrebbe essere formulata come segue: *ogni attività linguistico-cognitiva è calcolabile da una macchina di Turing*, (il che non vuol dire, ovviamente, che la nostra mente funziona *come* una macchina di Turing, ma che ogni attività mentale è simulabile da un dispositivo che abbia la stessa potenza computazionale di una macchina di

[34] M. Frixione (1994), p. 58.

[35] Ibidem.

[36] P. N. Johnson-Laird (1988), p. 58. Cfr. anche B. J. Copeland, *Computation*, in L. Floridi (2004), pp. 7-10.

Turing)»[37]. Di Francesco va un po' oltre la questione e afferma che «l'idea di Turing riguarda l'ontologia della mente: se anche l'intelligenza artificiale si rivelasse un programma completamente fallimentare, il modello computazionale non crollerebbe automaticamente con essa»[38], ovvero il reale interesse di Turing non sta nell'affermare, con atteggiamento comportamentista, che la macchina pensa se imita il comportamento umano bensì che «noi pensiamo perché il pensiero è una forma di computazione»[39], rivelando la propria attenzione all'uomo più che alla macchina.

I contributi forniti da Leibniz, Babbage e Turing vengono chiamati da Pratt tre «progetti», la cui importanza è da rintracciare rispettivamente nelle concettualizzazioni del meccanismo, della matematica e del ragionamento, ognuna delle quali tipiche del suo tempo. Infatti Pratt spiega: «L'obiettivo generale di Leibniz di costruire una macchina che ragioni è in qualche modo raggiunto dal "sistema esperto" [di cui si dirà più avanti] del periodo successivo a Turing, malgrado le indicazioni di Leibniz su come lo si potesse realizzare siano state respinte nei loro principi essenziali. Allo stesso modo, la nascita dei moderni calcolatori automatici ad alta velocità fu accolta, in un certo senso a ragione, come l'avverarsi del sogno di Babbage, nonostante la nostra concezione dei sistemi simbolici, e dell'algebra in particolare, abbia nel frattempo subito un cambiamento abissale. Possiamo capire che cosa Leibniz intendesse per "ragionamento" e che cosa fosse l'"algebra" per Babbage, e così pure cose intendesse Turing per "pensiero". Sebbene le nostre nozioni di tali concetti non siano più quelle che erano, possiamo riconoscere ogni pensatore come se stesse affrontando il problema di meccanizzare, completamente o in parte, i nostri processi mentali»[40]. Sarebbe opportuno qui considerare anche il contributo di Hilbert da cui prese avvio il lavoro di Turing. «La proposta di Hilbert è di considerare le regole logiche come procedure che si attivano in virtù della *forma* dei segni cui si applicano. Se poi consideriamo – come faceva Hilbert – le regole logiche come regole di pensiero ci si apre chiaramente la possibilità di considerare un processo di pensiero come un insieme di trasformazioni di simboli che si applicano alla

[37] M. Frixione (1994), p. 59. Se non fosse chiaro il riferimento al comportamentismo, Frixione precisa nella stessa pagina che «lo stesso test di Turing è stato a lungo considerato come il criterio più adeguato per stabilire la validità empirica dei modelli computazionali della mente. In realtà è ancora preistoria, ancora pesantemente condizionata da assunzioni di tipo comportamentista».

[38] M. Di Francesco (1996), p. 143.

[39] Ibidem.

[40] V. Pratt (1990), p. 12.

loro forma, cioè a una loro proprietà *materiale* di alto livello»[41]. Difatti il merito di Turing si intravede nel portare a concretezza la visione di Hilbert: «nessuno ancora pensava a un oggetto fisico reale. Erano macchine della mente, metodi o algoritmi per produrre risposte. Era come se si fosse concepita l'idea del software prima che esistesse un hardware su cui farlo girare. Anche se la macchina di Hilbert fosse esistita, non avrebbe avuto alcuna utilità pratica, perchè con ogni probabilità la quantità di tempo necessaria a tale macchina per stabilire se un'equazione qualsiasi avesse soluzioni sarebbe stata maggiore della vita dell'universo»[42].

Sotto l'influenza di Wiener, per cui «il sistema nervoso umano e i sistemi di comunicazione meccanici come la rete della *Bell Telephone* non sono distinguibili, almeno da un punto di vista teorico»[43], un gruppo di ricercatori, fra cui matematici e ingegneri, costituì quell'area di ricerca chiamata *Cibernetica*.

Dal greco[44] «κῦβερνητικός» ovvero «arte del timoniere, del pilota», alludendo all'unità di interessi che li accomunava ma soprattutto al «pilota che la nostra mente è rispetto al mondo e ai sui oggetti»[45], essa si occupava sostanzialmente di applicare l'«ingegneria dei controlli alla comprensione di processi fisiologici e neurofisiologici»[46]. Furono così realizzati i primi modelli su calcolatore, segnando l'inizio dell'uso delle macchine come simulatori.

Il passaggio dalla Cibernetica all'Intelligenza Artificiale avvenne sostanzialmente nel momento in cui il lavoro di Warren McCulloch e Walter Pitts del 1943 che, introducendo la logica nella cibernetica, aveva dato vita alla nozione del «neurone formale» e allo sviluppo delle reti neurali artificiali, fu messo in discussione da Marvin Minsky e Seymour Papert. Nel 1954 Minsky intuì la necessità di dotare le macchine della capacità di apprendimento ammettendo «una strutturazione del sistema iniziale molto più elaborata (sotto forma di simulazione mediante calcolatore digitale)»[47]. Infatti due furono i "cibernetici" che influenzarono l'I.A.: Shannon (ingegnere, poi professore) che, oltre lo sviluppo dei programmi di scacchi, confezionò la teoria dell'informazione, e Frank Rosenblatt che realizzò nel 1958 il

[41] M. Di Francesco (1996), p. 137.

[42] M. Du Sautoy (2004), p. 336.

[43] S. Williams (2003), p. 17.

[44] A sua volta ripreso dal sanscrito «kubara», timone (cfr. V. Di Spazio 2004).

[45] A. G. Biuso (2004), p. 27.

[46] V. Pratt (1990), p. 243.

[47] Ivi, p. 245.

perceptron, una macchina fondamentale per gli sviluppi della disciplina appena nata. Ma il prototipo di Rosenblatt era costituito da «puntatori» (*layer*) che mettevano in relazione gli input con i «neuroni logici» di McCulloch e Pitts, a loro volta collegati con puntatori che davano il risultato di uscita (output), sulla base della «legge di Hebb» riguardante l'apprendimento nelle cellule nervose (una connessione tra due neuroni tende a rinforzarsi se attivata a brevi intervalli di tempo); nel 1969 Minsky e Papert portarono alla luce le difficoltà che il *perceptron* dimostrava avere nel calcolare degli stimoli visivi, per cui era stato progettato.[48] Fu così decretata fallimentare la soluzione cibernetica al problema dell'intelligenza artificiale e l'interesse generale dei ricercatori passò presto dallo studio delle reti nervose allo studio dell'apprendimento, dividendo l'occidentale I.A. dalla sovietica Cibernetica. I principali esponenti dell'I.A. (o A.I. in lingua inglese) furono McCarthy, Minsky, Newell e Simon. Il primo, che coniò l'espressione in una conferenza storica al Dartmouth College in New Hampshire nel 1956, e il secondo stabilirono le loro ricerche al MIT e a Stanford, mentre il terzo e il quarto a Carnagie. Un terzo ricercatore che lavorò separatamente dagli altri due centri fu Selfridge che si interessò al riconoscimento delle forme da parte del calcolatore e alla capacità di inserire dati sotto forma di testo.

«Il programma che gode della reputazione di essere "la prima incursione della ricerca sull'Intelligenza Artificiale nei processi intellettivi di alto livello"»[49] fu quello curato, sin dai tempi del seminario di Dartmouth, da Newell e Simon giacché si proponeva che questo programma, chiamato *Logical Theorist* o LT, risolvesse problemi di logica. In Williams si legge l'entusiasmo di Simon per il suo LT, descritto dallo stesso ideatore come «un software capace di pensare in modo non numerico, e quindi di risolvere l'antico problema mente/corpo spiegando come un sistema composto di materia possa avere le proprietà della mente»[50] o esordendo ancora con frasi quali: «esistono ora macchine che pensano, imparano e creano»[51]. Inutile dire che usò termini descriventi capacità veramente fuori la portata reale della macchina. Il grande successo che ebbe il *Logical Theorist* nel dimostrare in tempi brevissimi 38 dei 52 teoremi del secondo capitolo dei *Principia Mathematica* di Russell e Whitehead (dopo aver ironicamente superato le prime difficoltà in termini di tempi millena-

[48] Cfr. R. Cordeschi, *Cybernetics*, in L. Floridi (2004), pp. 189-192.

[49] V. Pratt (1990), p. 268 che cita Feigenbaum nella nota 43 (p. 270) del suo testo.

[50] S. Williams (2003), p. 42.

[51] Ivi, pp. 42-43.

ri di esecuzione) fu considerato un vero e proprio fallimento. Se lo scopo dei progettisti del programma fosse stato quello di realizzarne uno velocissimo e con alte capacità risolutive, allora il LT sarebbe stato acclamato come un successo; ma non fu così, perché Newell e Simon avevano un'ambizione diversa ovvero riuscire a simulare l'abilità dimostrativa di un essere umano, con i suoi limiti caratteristici e con i suoi pregi. Questo comportò che il programma, con la sua altissima velocità d'esecuzione, risultò distante da una tipica esecuzione umana di dimostrazioni matematiche, portando a considerare il progetto come fallimentare. Ma «i matematici e i logici interessati ai fondamenti della matematica non avevano nessuno di questi interessi psicologici, ed essi sottolineavano […] che utilizzando algoritmi differenti si sarebbero potute generare dimostrazioni in modo più efficiente ed affidabile»[52], in riferimento a quelli usati da Newell a base *euristica* ovvero con la capacità di ridurre enormemente il numero di possibilità da verificare aumentando le probabilità di successo in tempi ristretti (anche se riducendone la garanzia).[53] Newell poi migliorerà LT in GPS (un *General Problem Solver*) con la caratteristica di creare per ogni obiettivo dell'algoritmo un sotto-obiettivo, e inoltre inventò «un programma chiamato traduttore, che convertiva i numeri in ingresso, espressi in notazione decimale, nella forma binaria d'obbligo, e che consentiva di usare codici mnemonici invece che numeri (ancora una volta traducendoli in forma binaria) per specificare le istruzioni»[54]. E' così possibile, con sotto-programmi che fanno riferimento direttamente all'*hardware* del calcolatore, programmare con simboli, decisi arbitrariamente dal programmatore, forniti in ingresso, rappresentanti operazioni precise eseguite dalla macchina: nacque l'idea del *linguaggio ad alto livello*.

I *sistemi esperti* sono il risultato dei sistemi a base euristica e, presentando una registrazione semplificata di conoscenze umane e di tecniche di ragionamento (cioè «basi di conoscenza») oltre un «motore inferenziale» che risolve problemi simulando il ragionamento euristico umano accedendo alla base di conoscenza fornita, riescono a trattare «problemi in ambito specialistico (ad esempio in un qualche settore diagnostico) con prestazioni, almeno nei casi riusciti, comparabili a quelle di un esperto umano del settore»[55]. Come anticipato, quello cibernetico è ritenuto oggi un esperimento che non ha condotto dove si era ipotizzato alla

[52] V. Pratt (1990), p. 277.

[53] Il nome di «euristica» si deve a G. Polya (1944). Cfr. S. Williams (2003), p. 17.

[54] V. Pratt (1990), p. 278.

[55] E. Burattini-R. Cordeschi (2001), p. 24.

sua nascita, ma ha rappresentato una fase importante dello sviluppo dell'I.A. Con i sistemi esperti era nata dunque la cosiddetta «I.A. simbolica» ovvero basata sulla conoscenza, da cui prende le distanze la «nuova I.A.» (di cui parleremo più avanti) ovvero quella ricerca sviluppatasi in più modi avente l'obiettivo di superare i limiti dell'I.A. simbolica.

Insieme a quello dell'Intelligenza Artificiale, intorno agli anni Cinquanta, si affermarono il programma delle scienze cognitive, in concomitanza alla psicologia cognitiva, e la linguistica detta "generativa" di Chomsky (che fu membro dello storico simposio tenutosi al MIT nel 1956). Come afferma Nannini, l'analogia tra mente e software e la linguistica chomskyana diedero vita al cognitivismo in psicologia e al funzionalismo in filosofia della mente.[56] Williams punta invece l'attenzione a ciò che la teoria chomskyana ha prodotto nel campo dell'I.A.: basandosi sui fattori biologici dell'apprendimento e dell'intelligenza, egli mina «alla base i tentativi della comunità dell'I.A. di costruire sistemi di linguaggio naturale *ex situ*»[57]. Il *cognitivismo* consiste essenzialmente nello «studio dei processi cognitivi al livello degli algoritmi»[58]; difatti per i cognitivisti l'attività mentale corrisponde a «un processo di elaborazione dell'informazione eseguito mediante la manipolazione di simboli in conformità con regole sintattiche»[59]. Anche se lontano dal dualismo, Chomsky si contrappone al comportamentismo di Skinner e avvalora la storica tesi dell'innatismo, risalente a Platone ma rintracciabile in Descartes e Leibniz, applicandola non soltanto alla capacità di apprendimento umano della lingua (tramite una "grammatica universale"), ma con la quale si desume che «anche tutte le altre attività umane che richiedono intelligenza siano governate dalle rispettive "strutture cognitive"»[60], tramite l'interazione del patrimonio innato (dunque lo stato interno) con l'ambiente esterno.

Il rapporto con le neuroscienze può diventare il terreno di scontro tra diverse posizioni concettuali in filosofia della mente, come per il cognitivismo e l'*eliminativismo*. Al contrario delle scienze cognitive, per cui la psicologia e la filosofia della mente debbano essere indipendenti dalle neuroscienze, l'eliminativismo (di Quine, Rorty, Stich e altri) suggerisce saldamente di sostituire tutti i termini riferiti al mentale «con un vocabolario attinto dalla

[56] Cfr. S. Nannini (2002), pp. 100-101.

[57] S. Williams (2003), p. 19.

[58] D. Marconi (2001), p. 12.

[59] S. Nannini (2002), p. 102.

[60] Ivi, p. 104.

neurobiologia»[61], radicalizzando il comportamentismo psicologico da cui nasce. Più precisamente, nell'ottica di Quine e Rorty ciò che va eliminato è «la *teoria* ontologica dualistica che fa da sfondo alla psicologia del senso comune»[62], e in particolare Rorty, ritenendo che «non ci siano affatto stati mentali né da distinguere né da identificare, totalmente o approssimativamente, con stati cerebrali»[63], imputa ad un errore culturalmente e storicamente condizionato il parlare «di sensazioni invece che di neuroni»[64]. In realtà va detto che Stich ha rivisto il proprio eliminativismo sostenendo che «sebbene rimanga vero che la filosofia del senso comune è profondamente errata, questo non giustifica la conclusione che gli stati mentali di senso comune – credenze, desideri, intenzioni – non esistano. Per passare dalle premesse critiche alla conclusione, è necessario considerare una serie di ulteriori passaggi, che dimostrano che no è così semplice eliminare gli stati mentali»[65]. Questa posizione che, come abbiamo visto, è stata rinvigorita recentemente dai coniugi Churchland, si inserisce tra quelle del materialismo, accanto alla *teoria dell'identità* (di Place, Smart e in parte Armstrong) e al *monismo anomalo* (di Davidson). Secondo Place «credere che la distinzione tra stati mentali e processi cerebrali provi l'esistenza di una qualche realtà immateriale»[66] è un'illusione perché non esistono "contenuti fenomenologici" ma solo stati cerebrali, che causano la cognizione di un'esperienza (parla di identità "contingente"). Si tratta di una precisa critica alla concezione brentaniana prima e husserliana poi dell'esperienza, infatti Place parla di «fallacia fenomenologica»[67]. Così continua Smart, sulla stessa linea, per cui «i termini psicologici non implicano di per sé l'esistenza di entità mentali; sono compatibili sia con il dualismo che con il materialismo, ossia sono 'topico-neutrali'»[68]. Con Armstrong diviene difficile separare nettamente la linea di demarcazione tra materialismo e funzionalismo, giacché è stato più difficile che in altri casi affermare con precisione il proprio materialismo, seppur netto, sostenendo infine che «gli stati mentali siano intrinsecamente ed essenzialmente degli stati cerebrali, che possono essere più facilmente (o anche esclusivamen-

[61] Ivi, p. 121.

[62] Ivi, p. 122.

[63] Ivi, p. 124.

[64] Ibidem.

[65] E. Carli (2003), p. 199.

[66] S. Nannini (2002), p. 107.

[67] Ibidem, ripresa dal testo dello stesso Place.

[68] Ivi, p. 109.

te) individuati mediante le funzioni che sono capaci di svolgere»[69]. Il materialismo di Davidson si può così riassumere: il linguaggio psicologico deve essere integrato con quello fisico nel senso che, non esistendo «alcuna legge psicofisica che correli eventi mentali tra loro simili a eventi fisici tra loro altrettanto simili»[70], le descrizioni psicologiche esplicano gli eventi mentali in un modo che, se ridotti a eventi fisici, non sarebbero comprensibili (questa è l'anomalia del mentale, cioè la compensazione con criteri esplicativi, in termini psicologici, dell'assenza di leggi strette tipiche del campo fisico), pur identificando ogni singolo evento mentale con un evento cerebrale. «In altre parole, l'anomalia del mentale impedisce la corrispondenza tra *tipi* di eventi mentali e *tipi* di eventi fisici, ma non impedisce quella tra loro *singole occorrenze*»[71]: posizione particolare e variamente interpretata. Nannini precisa che «l'ipotesi più ragionevole al riguardo sembra essere, comunque, quella che Davidson, pur non dichiarandosi mai fisicalista e tantomeno aderendo a quel 'fisicalismo delle occorrenze' che […] Fodor trarrà dai suoi scritti, sostenesse tuttavia negli anni Settanta un monismo di tipo sostanzialmente fisicalistico e che solo in tempi recenti, cambiando tacitamente posizione, abbia invece reinterpretato a favore del monismo neutrale la teoria formulata venti anni prima»[72].

Le critiche alle varie forme di materialismo non sono mancate ma non hanno tardato a sopraggiungere neanche le repliche dei materialisti: Nagel[73], riprendendo Kripke, e Jackson hanno criticato l'approccio materialista – cercando con difficoltà di distanziarsi al tempo stesso da dualismi di ogni sorta – differenziando l'esperienza soggettiva in prima persona dall'oggettività in terza persona, tipica della scienza. Le diverse modalità epistemiche (qualitativa e descrittiva) non sono però, obiettano i materialisti, diverse modalità ontologiche. La critica anti-materialistica, per cui «la differenza tra la conoscenza neurofisiologica del

[69] Ivi, p. 113 (cfr. anche nota annessa).

[70] Ivi, p. 124.

[71] Ivi, p. 125. «La distinzione fra tipo (*type*) e occorrenza (*token*) si riferisce alla differenza fra una classe di eventi (il tipo) e un membro particolare della classe (un'occorrenza)» (W. Bechtel 1992, pp. 157-158).

[72] S. Nannini (2002), p. 126.

[73] Ci riferiamo naturalmente all'antiriduzionista Thomas Nagel, e non al riduzionista classico Ernest Nagel per cui la riduzione è ancora una «relazione, dotata di una specifica struttura logica tra due *teorie* scientifiche» (M. Di Francesco 2000, p. 50).

cervello e il provare esperienze soggettive è profonda»[74], resta un tentativo, fondamentale e ancora per nulla invalidato, di evitare la riduzione fisicalista.

Sono famosi gli esperimenti mentali utilizzati per criticare il materialismo. Nagel, sfruttando l'argomento di Kripke per cui i *qualia* (sensazioni soggettive e private) non sono oggetti descrivibili distintamente dall'esperienza fenomenica proprio secondo la loro definizione, avanza un'idea che mette in luce l'impossibilità pratica di *essere un pipistrello*, ovvero di provarne la sensazione, proprio a causa dell'impenetrabilità e separatezza epistemica dell'esperienza fenomenica.[75] Jackson propone di riflettere su cosa saprebbe in più una esperta neuroscienziata (*Mary*) quando vede i colori rispetto a quando era chiusa in un laboratorio privo di colori: per l'autore il soggetto sperimenterebbe qualcosa di nuovo cioè «il quale» dei colori. I materialisti naturalmente distinguono il *saper fare* dal *conoscere* (distinzione, precisamente, fatta da Ryle) attribuendo al nuovo sapere di Mary un'esperienza del primo tipo e non del secondo.[76]

Alcune volte gli anti-riduzionisti sposano una concezione del rapporto mente-corpo e della coscienza come inintelligibili, aldilà della comprensione umana (questa è la posizione di McGinn di cui parleremo fra poco) o comunque come concetti velati di mistero ancora da chiarire; tra questi i più autorevoli sono Searle, Putnam e Nagel. Dei primi due si dirà più avanti diffusamente, mentre per quanto riguarda Nagel si può precisare subito che ha definito la propria posizione come *agnosticismo motivato*: pur ritenendo non impossibile che la filosofia o la scienza colmino il divario tra ignoto e noto, anche se in tempi lunghissimi, egli precisa che fino a quando la tradizionale analisi scientifica escluderà dal proprio campo d'indagine l'aspetto soggettivo, qualitativo, fenomenico della natura sarà incompleta e incapace di spiegare la coscienza.[77] Riguardo McGinn, in Di Francesco egli è definito, riprendendo l'analisi di Flanagan, *naturalista anticostruttivo* o *neomisterianista*.[78] Anche in Horgan c'è un richiamo ai «misteriani» di Flanagan,[79] infatti così vengono descritti da Horgan quegli studiosi e quegli scienziati che pensano, come Stent, che probabilmente il cervel-

[74] S. Nannini (2002), pp. 137-138.

[75] Cfr. T. Nagel, *Che cosa si prova ad essere un pipistrello?*, in D. R. Hofstadter-D. C. Dennett (2001), pp. 379-392.

[76] Cfr. F. Jackson, *What Mary didn't know*, in T. O'Connor-D. Robb (2003), pp. 458-463.

[77] Cfr. D. Marconi (2001), p. 24.

[78] Cfr. M. Di Francesco (1996), p. 206.

[79] Cfr. J. Horgan (1998), pp. 274 e ss.

lo potrebbe «non essere in grado di fornire una spiegazione di se stesso»[80]. In particolare, di McGinn viene precisata la dipendenza, scoperta dopo la formulazione del proprio, dal pensiero di Chomsky, per cui «la nostra capacità linguistica ci consente di formulare quesiti e di risolverli in modi che sono preclusi ai ratti, ma in ultima analisi anche noi ci troviamo a fronteggiare misteri altrettanto impenetrabili»[81]. Per McGinn, misteriano «forte» al contrario di Nagel misteriano «debole», «i grandi problemi filosofici sono reali, ma sono fuori dalla portata della nostra capacità cognitiva»[82], infatti afferma: «Abbiamo tentato per lungo tempo di risolvere il problema mente-corpo. Esso ha resistito tenacemente ai nostro migliori attacchi. Il mistero persiste. Penso sia giunta l'ora di ammettere tranquillamente che non siamo in grado di risolvere il mistero. Ma penso anche che questa stessa insolubilità – o la sua ragione – rimuove il problema»[83]; «i limiti della nostra mente non sono certo i limiti della realtà»[84]. Anche Fodor aderirà ai misteriani di cui parla Flanagan, ritenendo "misteri della mente" la coscienza e la natura dei processi mentali.[85]

E' utile porre sin da ora la distinzione tra due ambiti di ricerca nel campo dell'Intelligenza Artificiale che John Searle ha denominato *forte* e *debole*. Entrambe nascono intorno agli anni Cinquanta del secolo scorso ma si differenziano più che nella pratica negli obiettivi: mentre l'I.A. forte è quella, propriamente nata al congresso di Dartmouth del 1956, che si prefigge di dotare una macchina, dunque una struttura non protoplasmatica, di intelligenza umana (in qualsiasi modo la si voglia definire), di coscienza, autocoscienza e intenzionalità (posto che ogni ricercatore possiede la propria idea, spesso distinta dagli altri colleghi, di cosa sia utile per raggiungere tali obiettivi), l'I.A. debole è quella ricerca incentrata sul conseguimento di continue migliorie nel campo della simulazione di attività umane per mezzo di macchine (calcolatori digitali),. Più precisamente «il programma dell'I.A. *debole* ha un'impostazione sostanzialmente operativa e strumentale, rivolta a dotare l'umanità di po-

[80] Ivi, p. 291, dove riprende lo stesso Stent.

[81] Ivi, p. 237.

[82] Ivi, p. 97.

[83] C. McGinn, *Can we solve the mind-body problem?*, in T. O'Connor-D. Robb (2003), p. 438. Nel testo: «We have been trying for a long time to solve the mind-body problem. It has stubbornly resisted our best efforts. The mystery persists. I think the time has come to admit candidly that we cannot resolve the mystery. But I also think that this very insolubility – or the reason for it – removes the philosophical problem».

[84] Ivi, p. 452. Nel testo: «the limits of our minds are just not the limits of reality».

[85] Cfr. F. Cimatti (2004), p. 115.

tenti macchine in grado di svolgere con efficacia lavori non solo di routine e molto pesanti – cosa che esse hanno fatto egregiamente almeno a partire dalla Rivoluzione Industriale – ma anche di *simulare* attività intelligenti in quei settori nei quali non basta la semplice forza meccanica»[86]. E' dunque evidente che la seconda ha ottenuto un successo maggiore della prima, data la modestia e la concretezza degli obiettivi postisi.

E' giunto il momento di trattare (seppur non approfonditamente) l'elemento teorico, già più volte accennato, che ha come origine ma che, al tempo stesso, ha corroborato, nella sua interezza, gli studi sull'I.A.: il funzionalismo.

Il funzionalismo di cui qui si vogliono discutere le caratteristiche essenziali e gli autori fondamentali non è che una forma particolare di un concetto più ampio che abbraccia molte discipline.[87] Il funzionalismo in psicologia è stato introdotto da James e Dewey, con influssi di Brentano e della teoria evolutiva darwiniana, trattando i processi mentali come strategie che *fungono* alla sopravvivenza adattativa dell'organismo. Da questa interpretazione psicologica il funzionalismo di Putnam e altri prende spunto, ma con termini profondamente differenti e radicati in un ambito pur sempre psicologico ma tendente ad una prospettiva meno limitata che alla sola psicologia. Il funzionalismo psicologico infatti, per certi versi, è addirittura opposto a quello di Putnam: per il primo, «in rapporto all'ambiente, mente e corpo si comportano nello stesso modo e costituiscono di fatto un'unità inscindibile»[88], mentre per il secondo mente e corpo sono indipendenti, come vedremo meglio subito.

---

[86] A. G. Biuso (2004), p. 52.

[87] Il funzionalismo può essere riferito, prima che alla filosofia della mente e alla scienza cognitiva, ai seguenti ambiti: antropologia (ad opera di Malinowski e Radcliffe-Brown) e sociologia (a cura di Parsons, Merton, Davis, Hempel, Boudon e E. Nagel), per le quali le strutture sociali non hanno matrice storica o geografica ma funzionale alla società; linguistica (in particolare fonologia, morfologia, sintassi e semantica, nelle opere di Martinet, Jakobson, Saussure, Hjemslev e Schmidt), in cui si adotta l'identificazione dell'unità linguistica con la funzione che svolge; infine psicologia per cui i processi mentali sono prodotti adattativi dell'organismo.

[88] AA. VV. (1993), alla terza voce «funzionalismo», dove si precisa un aspetto importante della teoria, riguardo i processi mentali coscienti: «anziché concepirli come eventi psichici puri e irrelati col resto dell'organismo e con l'ambiente, [i funzionalisti] li interpretano come l'espressione massima dell'adattamento biologico; la coscienza compare ogni qual volta il comportamento automatico è ostacolato da eventi nuovi e imprevisti, che pongono problemi la cui soluzione non è contenuta nel repertorio del comportamento automatico stesso, ma che è indispensabile per la sopravivenza, a lungo o breve termine, dell'organismo biopsichico; una volta svolta la propria funzione adattiva, la coscienza si eclissa, lasciando al proprio posto i nuovi automatismi che essa stessa ha instaurato. Sul piano epistemologico, il funzionalismo [psicologico] si pone come

In questa seconda accezione, infatti, il funzionalismo si fonda sull'analogia, totalmente moderna, secondo la quale il cervello (o il corpo) *sta* alla mente *come* la parte materiale di un computer *sta* alla sua parte programmata (parti dette rispettivamente *hardware* e *software*).
Presupponendo la nozione kantiana di mente come processo e non come sostanza, oltre i risultati raggiunti qualche anno prima dagli psicologi americani G. Miller e U. Neisser,[89] nel 1960 Hilary Putnam, con l'articolo *Menti e macchine*, afferma che il rapporto tra mente e corpo è paragonabile a quello tra gli "stati logici" e gli "stati strutturali" di una macchina, e che dunque la mente è una macchina di Turing. Più precisamente afferma: «poiché sto sostenendo che esiste una *stretta analogia* tra il problema mente-corpo e quello delle relazioni tra stati logici e stati strutturali, ma non che i due problemi siano *identici*, tutto quello che mi occorre è che esista *un'analogia* adeguata tra il "linguaggio" di macchina ed il linguaggio umano»[90]. Afferma anche che «l'organizzazione funzionale (soluzione di problemi, pensiero) dell'essere umano o della macchina può essere descritta in termini di successione, rispettivamente, di stati mentali o di stati logici (nonché delle corrispondenti verbalizzazioni), senza fare alcun riferimento alla natura della "realizzazione fisica" di tali stati»[91].
Nel 1964 in *I robot: macchine o vita creata artificialmente?*, il filosofo radicalizza il carattere indipendente della struttura logica rispetto ad una realizzazione fisica, passando da una forma di materialismo ad una di dualismo: seppure neghi che «avere una psicologia» significhi essere coscienti, afferma che più robot possono avere la stessa psicologia (cioè processi mentali intesi verbalisticamente) anche avendo diverse costituzioni fisiche, per cui «se si intende il materialismo come la negazione dell'esistenza di attributi "non fisici", allora il

---

l'antitesi dell'elementismo introspezionistico, e in questa rivendicazione della priorità del "tutto" sulle "parti" o "elementi" precede il gestaltismo» e dunque anche l'olismo.

[89] I due psicologi cognitivi concepirono la mente come un «meccanismo centrale di elaborazione dell'informazione […] un complesso processo di elaborazione degli input sensoriali che li trasforma in output motori» (S. Nannini 2002, p. 101).

[90] H. Putnam, *Menti e macchine*, in H. Putnam (1987), p. 414. Anche Newell e Simon tentarono un approccio simile a quello di Putnam identificando un SFS (un «Sistema Fisico di Simboli») con la mente umana: «la tesi "forte" di Newell e Simon afferma che gli SFS, con le nozioni collegate di denotazione e interpretazione, forniscono una base esplicativa per tutto il comportamento intelligente degli esseri umani. Al giusto livello di analisi, le menti umane sarebbero identificabili con degli SFS, i quali sono dotati delle capacità di elaborazione simbolica *necessarie* e *sufficienti* per esibire intelligenza» (E. Burattini-R. Cordeschi 2001, p. 36).

[91] H. Putnam, *Menti e macchine*, in H. Putnam (1987), pag. 403.

materialismo è falso anche nel caso dei robot!»[92]. Comunque afferma sia che ritenere un robot cosciente è sì falso ma non contraddittorio, sia che esiste un parallelismo psicofisico tra uomo e macchina (cioè i robot sono «psicologicamente isomorfi»[93] agli uomini).

Nel 1973 Putnam rivede il proprio pensiero: *Filosofia e vita mentale* chiarisce che il concetto di isomorfismo funzionale, benché logicamente valido, si fonda sulla descrizione funzionale o psicologica dei sistemi in questione, ovvero un programma (che funzioni come una Macchina di Turing, per esempio) e la mente umana. Ma nonostante sia vero per Putnam che «due sistemi possono avere una composizione del tutto diversa ed essere funzionalmente isomorfi»[94] e «quale che sia il programma del cervello, deve essere fisicamente possibile (anche se non necessariamente realizzabile) produrre qualcosa che abbia quello stesso programma ma una composizione fisica e chimica completamente diversa»[95], egli afferma che gli stati psicologici non sono identici alle loro realizzazioni fisiche e soprattutto che l'identificazione della mente con una Macchina di Turing proposta nei suoi articoli precedenti è erronea, in base al fatto che uno stato psicologico umano è molto più complesso di uno stato logico di una MT.[96] Questo articolo presenta anche una forma di antiriduzionismo

---

[92] H. Putnam, *I robot: machine o vita creata artificialmente?*, in H. Putnam (1987), p. 423.

[93] Ivi, p. 426.

[94] H. Putnam, *Filosofia e vita mentale*, in H. Putnam (1987), p. 320.

[95] Ibidem.

[96] Più precisamente «uno *stato* di macchina di Turing viene descritto in modo tale che una macchina di Turing può trovarsi esattamente in uno stato alla volta. Inoltre, nel modello della macchina di Turing memoria e apprendimento non vengono rappresentati come acquisizione di nuovi stati, ma come acquisizione di nuove informazioni stampate sul nastro della macchina. Quindi, nell'ipotesi che gli esseri umani abbiano stati simili a quelli di una macchina di Turing, deve trattarsi di (1) stati in cui l'essere umano può trovarsi in qualsiasi momento, indipendentemente dalla memoria e dall'apprendimento, e (2) stati istantanei *totali* dell'essere umano, cioè stati che determinano, insieme con la memoria e l'apprendimento, quale sarà lo stato successivo, specificando altresì totalmente la condizione attuale dell'essere umano […]. Queste caratteristiche stabiliscono che *nessuno* stato psicologico, in qualsiasi senso ordinario del termine, può essere uno stato di una macchina di Turing. […] Fino a questo punto è ancora possibile che uno stato psicologico sia una lunga disgiunzione (in pratica una disgiunzione pressoché infinita) di stati di macchina […]. Ma questo diventa molto improbabile quando si passa da stati come "dolore" (che sono stati biologici) a stati come "gelosia", "amore" o "competitività". Essere geloso non è sicuramente uno stato *istantaneo*, e dipende da una gran quantità di informazioni e da molti fatti e abitudini appresi. Gli stati delle macchine di Turing sono invece istantanei e indipendenti dall'apprendimento e dalla memoria. […] la teoria è ora svuotata, giacché l'obiettivo originale [degli articoli precedenti] era di usare la tavola di transizione [di una MT] come modello di una teoria psicologica, laddove è

del mentale, per cui la scienza fisica e chimica non può spiegare il modo di funzionare della mente umana; il vero scopo, aggiunge Putnam, di una teoria del mentale deve essere dunque illustrare «la descrizione di questo livello autonomo del funzionamento della mente»[97]. A questo punto il filosofo ritrova il pensiero aristotelico e la formulazione della questione nei termini di materia e forma. Come dice Nannini, nel pensiero antico e in quello odierno le vie filosofiche sono percorse in modo parallelo: come Aristotele trovò la "terza via" tra Platone e Democrito, così Putnam ritrova Aristotele tra dualismo e materialismo.[98] Qualunque sia la sostanza di cui siamo fatti, la «forma intellettuale» è quella che interessa anche se è prematura ogni speculazione in merito quando ancora non si possiede nemmeno «una descrizione *idealizzata* dell'organizzazione funzionale del cervello»[99].

In saggi successivi Putnam ha addirittura criticato in maniera completa ogni residuo di funzionalismo presente nel suo pensiero esposto negli articoli fin qui citati. Infatti, nei primi aveva proposto il funzionalismo, poi nel 1973 ne ha corretto alcuni aspetti, e successivamente approda ad un *realismo interno* che poco o nulla condivide col funzionalismo (anche quello rivisto): non solo la mente non è una MT, ma gli stati mentali non sono più nemmeno riducibili a stati funzionali.[100] Secondo il principio hansoniano della *theory ladenness*,[101]

---

ora chiaro che la descrizione in termini di tavola di transizione, benché diversa dalla descrizione al livello di particelle elementari, è altrettanto lontana dalla descrizione mediante una teoria psicologica quanto la descrizione al livello fisico-chimico» (Ivi, pp. 325-326). Riguardo alla *tavola di transizione* Putnam spiega che «dire che un automa finito ha una certa tavola di transizione significa dire che *esistono proprietà* […] che l'oggetto possiede (nel senso, cioè, che in ogni istante ne possiede una), e che si susseguono l'una all'altra secondo una certa regola» (H. Putnam 1993a, pp. 338-339).

[97] H. Putnam, *Filosofia e vita mentale*, in H. Putnam (1987), p. 325.

[98] Cfr. S. Nannini (2002), pp. 119-120.

[99] H. Putnam, *Filosofia e vita mentale*, in H. Putnam (1987), p. 330. Interessante notare, come fa Nannini (cfr. S. Nannini 2002, p. 120), che Putnam ha congiunto il pensiero di Aristotele a quello kantiano, come fece anche Hegel: l'Idea è forma spirituale aristotelica e attività kantiana, processo. L'approdo hegeliano è distinto da quello del filosofo contemporaneo solo nella sua formulazione finale ma quanto alle fondamenta è parallelo ad esso, infatti anche per Putnam la mente è autonoma dal fisico (pur non cadendo nel dualismo) ed è un processo, un susseguirsi di attività.

[100] Ci riferiamo in particolare a H. Putnam (1993b), la cui edizione originale risale al 1988.

[101] Cfr. G. Boniolo-P. Vidali (2003), pp. 31 e ss.

per Putnam il mondo non esiste indipendentemente dalla cornice teorica entro la quale è descritto; il filosofo sconfinerà così nel relativismo.[102]

Molti filosofi e psicologi si sono accostati al funzionalismo del "primo" Putnam, condividendone l'assunto fondamentale (gli stati mentali possono essere definiti nei termini delle loro relazioni reciproche) ma differenziandosi nel «modo in cui tali interazioni devono essere specificate»[103].

Intervenendo sulla formulazione originaria del funzionalismo, Fodor e Block[104] hanno dato vita al funzionalismo detto *computazionale* che, considerando la mente «un'esecuzione di operazioni formali effettuate su simboli codificati al suo interno»[105], consiste nell'identificare uno stato mentale con un'operazione (un calcolo) che potrebbe essere eseguita da macchine diverse. Nei termini di Dennett, la mente è così «un motore sintattico che simula un motore semantico»[106] in quanto il filosofo ritiene valida l'affermazione di Haugeland per cui «bisogna occuparsi della sintassi (le operazioni formali), che la semantica si occuperà di se stessa»[107]. Della posizione di Fodor si è già accennato per quanto riguarda l'innatismo e il rapporto mente-corpo (fisicalismo delle occorrenze). Si può precisare che per Fodor esiste dunque un linguaggio del pensiero (il «mentalese») che, come un codice innato di matrice kantiana, interpreta e "calcola" l'informazione in entrata. Infatti gli stati intenzionali sono "scritti" nel cervello sotto forma di simboli combinabili in base a regole sintattiche. Perciò qualsiasi attività mentale è hobbesianamente una computazione per mezzo di regole sintattiche. Eppure Fodor tiene a precisare che la sua teoria computazionale e rappresentazionale della mente non è che una teoria psicologica e le sue implicazioni nel

---

[102] Cfr. S. Nannini (2002), pp. 144-145.

[103] W. Bechtel (1992), p. 188.

[104] Block ha criticato il funzionalismo per proporne una forma più efficace (come hanno fatto quasi tutti i funzionalisti) insieme a Fodor, ma sono risultate essenziali le obiezioni (la *pars destruens* della sua teoria): in particolare quella secondo cui il funzionalismo, se seguito nei suoi assunti, concederebbe il carattere di coscienza e funzionalità mentale a strutture che non possono averle, come nel caso ipotetico dell'intera popolazione cinese che sarebbe funzionalmente equivalente ad una mente umana se ogni cinese svolgesse un compito funzionalmente equivalente a quello svolto da una macchina di Turing (o da un processo mentale). Ma ciò è assurdo, facendo seguito così alle critiche di Nagel e Jackson prima trattate. Cfr. D. Marconi (2001), p. 50.

[105] W. Bechtel (1992), p. 194.

[106] Ibidem

[107] B. P. McLaughlin, *Computationalism, connectionism, and the philosophy of mind*, in L. Floridi (2004), p. 140. Nel testo: «Take care of the syntax (the formal operations), and the semantics will take care of itself».

campo dell'I.A. sono talvolta inappropriate: considerando la psicologia una scienza, Fodor afferma infatti che «l'intelligenza artificiale, anche per sua stessa ammissione, riguarda una sfera che non è quella della scienza – e certamente non è quella della filosofia – bensì quella dell'ingegneria»[108], pur ritenendo la propria teoria «un frammento di una psicologia cognitiva completa e soddisfacente»[109]. Aggiunge anche che la sua teoria non può dire nulla riguardo la coscienza, a differenza di altri computazionalisti.[110]

Sia Fodor che Dennett condividono l'analogia mente-software. Ma Dennett, neo-comportamentista riguardo all'intenzionalità delle macchine,[111] propone un funzionalismo detto *omuncolare* che, in termini funzionalisti, punta l'attenzione più alla struttura gerarchica di un sistema che al suo compito complessivo: presuppone che la struttura del mentale sia gerarchicamente costituita da *homunculi* capaci di svolgere delle attività elementari tali che se sommate a quelle degli altri possono formare gli stati mentali veri e propri.[112] Infatti Dennett concepisce il rapporto tra mente e corpo come una questione di "livelli". La *società della mente* minskyana (per cui la coscienza è «semplicemente un tipo di memoria a breve termine, un sistema di basso livello per conservare le registrazioni»[113]) e la cultura informatica dei linguaggi di programmazione (che suddivide la struttura di un software in base alla complessità dei livelli più superficiali di accesso ai dati, costruiti su livelli più semplici fino a giungere al *linguaggio macchina*, il più elementare dei livelli, seppur ancora distante dalla differenza logica di «0» e «1») hanno consentito a Dennett di formulare la sua concezione del rapporto tra cervello e mente.[114] Infatti il filosofo identifica gli omuncoli della mente di livello più basso con le operazioni del cervello stesso.[115]

---

[108] J. A. Fodor (2001), p. 3. Non è casuale che questa affermazione sia diametralmente opposta a quella di Biuso, citata nell'introduzione.

[109] Ibidem.

[110] Cfr. Ibidem (in nota).

[111] Ritiene infatti che «i "sistemi intenzionali" […] non necessariamente hanno effettivamente desideri e credenze; per considerarli 'intenzionali' è sufficiente che il loro comportamento divenga più facilmente spiegabile qualora si presuma che essi si comportino come se li avessero» (S. Nannini 2002, p. 171).

[112] W. Bechtel (1992), pp. 199-200.

[113] J. Horgan (1998), p. 283.

[114] Cfr. M. Minsky (1989) e D. R. Hofstadter-D. C. Dennett (2001).

[115] Per Dennett «non esiste un singolo flusso di coscienza unitaria, in quanto non esiste una versione preconfezionata dei nostri stati coscienti che *preceda* le procedure di accesso che li renderebbe espliciti. […] L'*illusione* di una continuità e coerenza di questo soggetto virtuale deriverebbe da una particolarità della no-

Le implicazioni nell'I.A. della teoria di Dennett si rintracciano nelle reti artificiali di Rumelhart e McClelland: il PDP (Parallel Distributed Processing).

A distanza di un decennio dal lavoro di Minsky e Papert del 1969, molti ricercatori ripresero a occuparsi di reti artificiali rifacendosi al connessionismo di Rosenblatt. Rumelhart e McClelland mostrarono infatti che «processi paralleli distribuiti» (PDP), reti «multi-strato» (*multilayer*) che associavano più unità «nascoste» agli input e agli output, con algoritmi che potevano "apprendere" dagli errori (*back-propagation*) erano in grado di superare le difficoltà notate da Minsky e Papert nel *Perceptron*.[116]

Oltre ai processi distribuiti, furono tentate altre vie per superare i vecchi modelli. Il connessionismo si è impegnato a superare le difficoltà relative a quelle capacità tipiche dell'intelligenza naturale sia di risolvere problemi con metodi non descrivibili a livello conscio (la "conoscenza tacita") sia di sapersi adattare a delle condizioni impreviste dello stato del mondo in rapporto al proprio stato interno (detta "robustezza"), basandosi essenzialmente sull'ipotesi per cui «i processi cognitivi possono essere simulati efficacemente solo da una macchina la cui struttura e il cui funzionamento siano simili a quelli del cervello: per l'elaborazione parallela anziché strettamente sequenziale, per il grande numero di connessioni tra unità "neuronali", per la semplicità dell'informazione trasmessa (attivazione o inibizione, varianti in modo continuo e non discreto come nei calcolatori digitali)»[117]. Insomma il connessionismo tenta di avvicinare il funzionamento di un calcolatore al funzionamento del cervello, considerando la peculiare struttura reticolare dei collegamenti tra neuroni (le cellule nervose) nel cervello: invece di analizzare i dati uno per volta, essi vengono processati contemporaneamente da più processori che rappresentano una struttura neurale; e soprattutto si intende superare la semplicità dei calcolatori non connessionistici in cui il segnale è interpretato in due soli modi (accesso/spento) fornendo al sistema una continuità del segnale in modo tale da poter interpretare questo in più di due modi (dunque in modo non discreto). Una *rete neurale* è un insieme di *nodi* collegati da *connessioni* «capaci di trasmettere *attivazione* o *inibizione* da un nodo agli altri. Ciascuna connessione è caratterizzata da un *peso*, che determina la misura dell'attivazione (o dell'inibizione) trasmessa. L'insieme

---

stra specie, quella di "tessere", "filare", ma anche "secernere" [*to spin*] un sé, capacità paragonabile […] a quella con cui il ragno tesse la propria tela o il castoro edifica la propria diga» (M. Di Francesco 1996, p. 230).

[116] Cfr. R. Cordeschi, *Cybernetics*, in L. Floridi (2004), pp. 191-192. Cfr. inoltre S. Rose (1994), pp. 105-109.

[117] AA. VV. (1993), alla voce «connessionismo».

delle attivazioni e inibizioni ricevute determina lo *stato di attivazione* di un nodo, e quindi la sua influenza sui nodi con cui esso è connesso in uscita. In una rete, un insieme di nodi (*unità di input*) è caratterizzato dal fatto di avere solo connessioni in uscita; un altro insieme (*unità di output*) dal fatto di avere solo connessioni in entrata. Normalmente, una rete ha anche altri nodi (tra quelli di input e quelli di output), detti *unità nascoste*. […] La rete ha "appreso" quando è in grado di dare risposte adeguate anche a input nuovi»[118].

Su queste base sono stati sviluppati gli algoritmi detti *genetici*, a opera di P. C. Holland. Si tratta di algoritmi che riflettono, nella loro struttura, la caratteristica dell'evoluzione naturale, cioè l'adattività che Darwin chiamava *fitness*.[119]

Modificazioni di questo tipo di algoritmi sono state utilizzate da Gerald Edelman per avvalorare a livello ingegneristico la propria teoria neurobiologica.[120] Per trattare le prime, è necessario conoscere per vie generali la seconda, resa pubblica prima in *Darwinismo neurale* (1987) e *Topobiologia* (1988), poi in opere quali *Il presente ricordato* (1989), *Sulla materia della mente* (1992) e *Più grande del cielo* (2004).

Edelman ritiene di appartenere al «circolo dei realisti» ovvero coloro che, come lui, rifiutano l'analogia mente-software funzionalista e cercano una convergenza tra la psicologia cognitiva e le empiriche neuroscienze.[121] Critica le teorie di Turing, Wolpert, Oster e altri affermando che «ciascuna di esse è incompleta, o presenta un certo numero di difficoltà

---

[118] Ibidem.

[119] Questi algoritmi «agiscono su una opportuna popolazione di *solutori* ad un determinato problema, ciascuno dei quali è codificato tramite una apposita stringa di simboli. Ogni solutore può essere caratterizzato quantitativamente tramite la sua *fitness*, cioè la sua capacità di risolvere il problema in questione. Ad ogni generazione si eliminano i solutori caratterizzati da una fitness troppo bassa e, contemporaneamente, si creano nuovi solutori tramite due meccanismi (agenti su stringhe di simboli) che garantiscono la variabilità genetica, noti come *mutazione* e *cross-over*. L'evoluzione della popolazione di solutori si arresta quando le prestazioni del miglior individuo raggiungono un livello accettabile» (E. Burattini-R. Cordeschi 2001, p. 114). In robotica evolutiva il procedimento è similare: degli agenti vengono dotati di sistemi di controllo, per esempio sensoriali, casualmente diversi, e successivamente vengono fatti «riprodurre» gli agenti il cui comportamento approssima di più quello voluto. Così si otterrà almeno un individuo in grado di produrre il comportamento desiderato. Cfr. Ivi, p. 130.

[120] Molti neuroscienziati si sono dedicati allo studio delle basi biologiche della coscienza ma, come afferma Benedetto Farina, solo Edelman ha «formulato una teoria complessiva che raccolga i dati sperimentali e che guidi le speculazioni future» (B. Farina 1999, pp. 133-134).

[121] G. M. Edelman (1993b), p. 355.

quando viene messa a confronto con le varie osservazioni evoluzionistiche, embriologiche e genetiche»[122]. Nella pubblicazione del 1988 afferma che «la topobiologia si occupa delle interazioni fra le superfici cellulari e tra queste e i substrati. Tali interazioni dipendono dalla sede in cui hanno luogo e hanno l'effetto di modificare la regolazione cellulare»[123]. Punto di partenza dell'intera ricerca sembra essere il fatto che «la risposta alla ubicazione di una cellula dipende da quando, nell'ambito di un processo storico, essa entra in contatto con certe cellule vicine. […] l'indirizzo esatto di una singola cellula non è il determinante fondamentale. Soltanto la posizione relativa di una cellula in un collettivo entro un certo periodo di tempo è importante»[124]. Presupponendo che «i geni specificano le regole locali, non la struttura globale»[125], le molecole che potrebbero candidarsi per la *regolazione topobiologica* (regolazione molecolare della *forma* degli organismi viventi) sono quelle con la caratteristica dell'«adesione» delle cellule fra loro e ai substrati. Esse sono sostanzialmente tre famiglie di molecole, appunto detto «di adesione»: CAM (*cell adhesion molecole*) – adesione fra cellule –, SAM (*substrate adhesion molecole*) – adesione al substrato – e CJM (*cell junctional molecules*) - «giunzione» fra cellule. Fondamentale è che «tutte queste cellule svolgono sia funzioni meccanochimiche sia di regolazione nel corso dello sviluppo»[126].

---

[122] G. M. Edelman (1993a), p. 29.

[123] Ivi, p. 9.

[124] Ivi, p. 38.

[125] Ivi, p. 72.

[126] Ivi, p. 149. Le CAM sono proteine «intrinseche alla membrana» (Ivi, p. 110) e possiedono due meccanismi per modificare il modo di legarsi o modificare i confini che si formano tra *collettivi di cellule*, cioè cellule raccolte in gruppi o meglio «gruppo di cellule adiacenti con proprietà fenotipiche comuni, che agiscono come una fonte di segnali o un loro bersaglio» (Ibidem). Il primo è l'«espressione» di una nuova CAM sulla superficie con tipo di legame diverso da quello delle cellule circostanti; il secondo è un cambiamento del modo di espressione della CAM sulla superficie della cellula, tramite «cambiamento, nel tempo, della quantità, della distribuzione, o delle proprietà chimiche di un tipo particolare di molecola a livello della superficie cellulare» (Ivi, p. 114). Infatti le CAM variano nella quantità, nella posizione o nella distribuzione, o nella struttura molecolare. E' così che i processi modificano il legame della CAM «influenzando il modo in cui i collettivi di cellule cambiano morfologia e segnalano l'uno all'altro le rispettive condizioni» (Ivi, p. 117). Dunque le CAM trasformano cellule isolate, in una determinata posizione, in collettivi multicellulari «con proprietà di sviluppo diverse da quelle di un gruppo di cellule isolate [, e sono responsabili] delle modificazioni dei processi di movimento, divisione e morte, cioè delle forze motrici che regolano la distribuzione delle cellule durante lo sviluppo» (Ibidem). Fondamentale è l'interazione dipendente dal contesto tra segnali locali e i collettivi tenuti insieme dalle CAM, poiché è questa interazione (ricezione e invio di segnali tra CAM e contesto)

Il fatto che «gli eventi dello sviluppo hanno un carattere epigenetico e sono controllati topobiologicamente»[127] e il fatto che «è la configurazione, e non semplicemente il differenziamento delle cellule, la base evolutiva della morfogenesi»[128] conducono direttamente all'ipotesi edelmaniana dei «morforegolatori», la cui «via della regolazione consiste di segnali chimici e meccanochimici (responsabili di una modulazione globale), che, insieme, attivano i geni morforegolatori per le CAM»[129], innescando l'attivazione anche delle SAM e CJM.

Mettendo a frutto le scoperte fatte nel campo dell'immunologia, valse il premio Nobel nel 1972, Gerald Edelman ipotizza, ma con prove sperimentali confortanti, che «la selezione naturale, durante l'evoluzione, lascerebbe sopravvivere soltanto quegli organismi in cui particolari combinazioni individuali di movimenti morfogenetici diversi e di espressioni diverse dei geni per le CAM portano ad adeguate sequenze induttive funzionali»[130]. Sulla scorta di quanto finora premesso, Edelman propone un «darwinismo neuronale» o «TNGS» (*the-*

---

che svolge la funzione principale nella morfogenesi. Il ruolo delle CAM non si ferma alle fasi precoci dell'embriogenesi ma continua negli eventi *istogenetici* successivi (tessuti complessi quali il sistema nervoso). Cfr. Ivi, p. 118. «Le CAM regolano la morfologia e sono regolate da essa: la distruzione del loro legame modifica la morfologia, e la distruzione della morfologia modifica la regolazione dell'espressione delle CAM. […] le CAM giocano un ruolo topobiologico fondamentale nello sviluppo: la loro struttura, funzione, e localizzazione a livello della superficie della cellula, la loro modulazione in sedi chiaramente distinte, secondo schemi completamente riproducibili, e la loro connessione con sequenze di segnalazione e differenziamento durante l'induzione e l'istogenesi giustificano ampiamente l'affermazione che le CAM sono molecole morforegolatrici fondamentali» (Ivi, p. 125). Le SAM compongono in buona parte la matrice extracellulare (ECM, *extracellular matrix*), membrana basale, e sono in grado di svolgere azioni coordinate con le CAM ma il loro nesso preciso è ancora oggetto di studi. «L'ECM può influenzare il movimento delle cellule, la mitosi, la forma e l'espressione genica. […] i componenti dell'ECM [sono] fondamentali elementi morforegolatori» (Ivi, p. 140). Le CJM sono molecole, differenti da CAM e SAM, che formano «complessi giunzionali specializzati» (Ivi, p. 141) che includono «giunzioni intervallate», «giunzioni strette», «giunzioni aderenti» e infine «desmosomi». Le prime realizzano una comunicazione chimica e elettrica tra cellule, le seconde saldano superfici dell'epitelio (cellule unite per mezzo di CAM, con polarità definita), le terze e gli ultimi collegano regioni di contatto tra cellule con elementi del citoscheletro, «un complesso assemblaggio di diverse proteine che influenza le interazioni in ogni parte della cellula» (Ivi, p. 251).

[127] Ivi, p. 155. Nel testo, in corsivo.

[128] Ivi, p. 156. Nel testo, in corsivo.

[129] Ivi, p. 179.

[130] Ivi, p. 195.

ory of neural groups selection). Il neuroscienziato considera che «persino il genoma umano, relativamente grande, non possiede che una frazione del DNA che sarebbe direttamente richiesta per definire le connessioni del cervello»[131] e i cambiamenti delle sinapsi durante il comportamento e lo sviluppo rendono ancora più «assurdi» i requisiti. In virtù di tale complessità cerebrale, egli critica ogni analogia tra cervello e computer perché, entrando nel dettaglio della struttura biologica, «l'analogia si interrompe, e ci troviamo di fronte a una serie di problemi le cui reciproche implicazioni equivalgono per le neuroscienze a una serie di crisi interpretative»[132]; infatti va inoltre considerato che «i neuroni stessi possono disporsi anatomicamente secondo una molteplicità di modi complessi, ma risultano talvolta disposti in mappe, un principio, questo, molto importante nei cervelli complessi»[133]. Culmine del lavoro edelmaniano è la nozione di *rientro*. Durante lo sviluppo del cervello, si formano connessioni ramificate che variano da individuo a individuo e per mezzo della selezione (che agisce come rinforzo sinaptica, estensione e contrazione dei neuriti, migrazione e morte di cellule) gruppi di neuroni si organizzano in lamine o nuclei del cervello. Poi «ha luogo un secondo processo di selezione durante il soggiorno dell'animale nel suo ambiente dopo un certo stadio dello sviluppo»[134], in cui la rete non è modificata ma «i gruppi che rispon-

[131] Ivi, p. 208.

[132] Ivi, p. 214. Suddette *crisi* possono essere *strutturali* se legate all'anatomia e allo sviluppo, difatti «la rete del cervello viene costruita [a differenza delle reti artificiali] durante lo sviluppo a seguito dei movimenti, delle estensioni e delle connessioni cellulari di un numero crescente di neuroni. Si tratta di un esempio di sistema auto-organizzato» (Ivi, p. 215); considerando infatti che «parti del cervello (in verità la maggior parte del suo tessuto) ricevono input solo da altre parti, e mandano output ad altre parti ancora, senza interventi dall'esterno» (Ivi, p. 209), certi principi dello sviluppo, che portano ad una variabilità continua delle connessioni e ramificazioni sovrapponentesi con configurazioni di sinapsi non identificabili e non necessariamente ripetibili, «mettono in crisi coloro i quali credono che il sistema nervoso sia preciso e "cablato" esattamente (come nei collegamenti interni del calcolatore)» (Ivi, p. 215). A questo proposito Edelman tiene a precisare che «non esistono due sistemi nervosi, neanche quelli dei gemelli, che siano uguali a livello della loro struttura fine» (Ivi, p. 238). Oppure si tratta di crisi *funzionali* se riguardano la fisiologia e la psicologia: la maggior parte delle sinapsi della rete nervosa sembra inattiva (*sinapsi silenti*), rendendo ancora misteriosa la loro funzione. Inoltre il fatto che certe mappe subiscano modifiche temporanee nei loro confini è un problema giacché non rappresentabile via software, essendosi constatato che «la variabilità delle mappe non è discreta o a due valori, ma continua, a grana fine, e, allo stesso tempo, distribuita» (Ivi, p. 217). Cfr. G. M. Edelman (1993b), p. 53 e pp. 345-352 (sul funzionalismo "alla Turing", in particolare).

[133] G. M. Edelman (1993a), p. 210.

[134] Ivi, p. 225. Nel testo, in corsivo.

dono meglio a un dato input vengono selezionati attraverso un aumento dell'efficacia delle loro connessioni sinaptiche»[135]. I gruppi selezionati, unità funzionali, per via *degenerativa* (cioè selettiva per soppressione del meno adatto, o *selezione negativa*) possono ora «rispondere a nuovi stimoli che sono simili, ma non identici, allo stimolo originale»[136]. Il «rientro» spiega come «un animale può generalizzare a partire da un piccolo campione»[137]. I neuroni, agendo come rilevatori di caratteristiche, «mappano […] in alcune lamine altamente organizzate nel cervello (mappa 1). Altri neuroni (per esempio quelli collegati al tatto leggero su un dito che so muove) agiscono come correlatori (muovendosi, per esempio, sui contorni di un oggetto)»[138] e mappano così in un'altra lamina (mappa 2). Ora, grazie a «connessioni rientranti» le due mappe «mappano l'una sull'altra […] cosicché i gruppi di una mappa possono eccitare i gruppi di un'altra»[139]. Così si ha un collegamento parallelo e indipendente tra caratteristiche (mappa1) e correlazioni di caratteristiche (mappa 2). Questa «coppia di classificazioni» è solo esemplare in quanto il cervello impiega più di due canali di comportamento, anzi un numero straordinario. Difatti la proprietà essenziale dei cervelli superiori è il loro essere «mappati in modo rientrante, in una misura che non si osserva in nessun altro oggetto materiale conosciuto»[140]: «in ogni momento la materia della mente interagisce con se stessa»[141]. L'intento della teoria del premio Nobel è essenzialmente diretto a «comprendere i fondamenti biologici della percezione»[142]; infatti egli tenta di spiegare «quali meccanismi neurali consentono la categorizzazione percettiva e la generalizzazione»[143]. Quindi la segnalazione rientrante, insieme alla selezione durante lo sviluppo e durante l'esperienza, è

[135] Ibidem. Nel testo, in corsivo.

[136] Ivi, p. 226.

[137] Ivi, p. 231. Alcuni critici (Marvin Minsky e Daniel Dennett soprattutto) di Edelman espressero il parere che il rientro non fosse altro che la «retroazione» o *feedback* su cui si basavano le strutture cibernetiche, ma Edelman confermò la distinzione dei due concetti: «*non* si tratta di retroazione, che si produce tra due linee, in cui si ha una funzione definita, un'istruzione» (J. Horgan 1998, p. 260).

[138] G. M. Edelman (1993a), p. 232.

[139] Ibidem.

[140] Ivi, p. 235.

[141] G. M. Edelman (1993b), p. 55.

[142] G. M. Edelman (1995), p. XVII.

[143] Ibidem.

il meccanismo responsabile del comportamento adattivo del sistema nervoso.[144] In questo modo Edelman può affermare come ineluttabile la «crisi dell'homunculus», di cui si è avuto modo di parlare a proposito di Daniel Dennett: «il cervello contiene una molteplicità di mappe che interagiscono senza alcun supervisore e che tuttavia rendono unitaria e coesa la scena percettiva»[145]. Per Edelman è infatti necessario seguire la via darwiniana: «il problema […] è analogo a quello che dovette affrontare Darwin quando studiò le origini dell'ordine tassonomico e l'origine delle specie senza presumere un intervento divino, ossia senza supporre l'esistenza di un superhomunculus»[146].

Passando agli aspetti più ingegneristici delle teorie di Edelman, va premesso che l'elemento essenziale della teoria è quello dei *valori*: «i valori sono, secondo il neuroscienziato, dei veri e propri indici di valore evoluzionisticamente determinati che la mente associa alle categorie e che rispondono a criteri che aumentano l'adattamento della specie»[147]. *Darwin I* fu il primo robot costruito sulla scorta delle idee sviluppate nella teoria: presentava un «generatore di repertori» (stringhe di bit casuali), un «generatore di stimoli» (stringhe di bit di input da accoppiare), una «regola di accoppiamento» (che ricercava le corrispondenze tra uno stimolo selezionato casuale e i singoli elementi del repertorio), e un «amplificatore» (che aumentava la probabilità che gli elementi del repertorio accoppiatisi con successo agli input rispondessero nuovamente in futuro).[148] *Darwin II* rispecchia più dell'automa precedente la struttura delle reti artificiali sfruttando il sistema della «coppia di

[144] Cfr. Prefazione di G. Tononi, in Ivi, p. XXV. Nel contesto di queste nozioni fondamentali si inserisce anche la trattazione della *memoria* intesa come *ricategorizzazione*: «la memoria, in un sistema selettivo, non è un deposito di dati che descrivono gli oggetti o gli eventi incontrati. E' invece un processo dinamico, il cambiamento nel modo di funzionare del cervello e dell'intero organismo a seguito di innumerevoli eventi selettivi precedenti. Grazie al rientro, la categorizzazione operata in parallelo da un grandissimo numero di gruppi neuronali è infatti costantemente attiva, sia durante la percezione che durante l'immaginazione. La memoria è quindi ricategorizzazione e non registrazione» (Ivi, p. XXVII). Cfr. anche G. M. Edelman (1993b), pp. 156-172. L'effetto che produce la ricategorizzazione «è proprio ciò che Edelman chiama il presente ricordato: ovvero un circuito di ricorsivo rientro tra il presente e se stesso e che dà il senso di presente rispetto a quello passato» (B. Farina 1999, p. 135).

[145] G. M. Edelman (1993b), pp. 113-114.

[146] G. M. Edelman (1993a), p. 218.

[147] B. Farina (1999), p. 135.

[148] Cfr. G. M. Edelman (1995), p. 59.

classificazione» di cui si è detto prima.[149] *Darwin III* è dotato di un occhio mobile e di un braccio con quattro articolazioni, il cui ambiente è simulato al calcolatore: inizialmente si muove a caso, poi con l'esperienza «eventi selettivi nel sistema nervoso di Darwin III producono numerose modificazioni sinaptiche, guidate, […] dai sistemi di valore»[150] che permettono all'automa di riconoscere oggetti per mezzo di «mappaggi globali». *Darwin IV* ha un «cervello» simulato «collegato ad un fenotipo reale»[151], è dotato di telecamera, sensori ad infrarossi, sensori di conduttività elettrica, magneti e ruote, libero di muoversi in un ambiente reale. Il cervello, simulato al computer, «opera secondo principi selezionasti» adattandosi bene a circostanze variabili: Darwin IV non è programmato esplicitamente, ma solo grazie a «modificazioni selettive di riflessi innati e comportamenti appresi concatenate e guidate dai sistemi di valore»[152]; lo definisce «la prima cosa non vivente davvero capace di imparare»[153]. Dunque Edelman utilizza i calcolatori «per *simulare* parti del cervello e anche per favorire la costruzione di macchine percettive che si basino sulla selezione, non su istruzione»[154] come i vari «organismi euristici» appena discussi.

Come il lavoro di Edelman, anche la *nuova robotica* e la *cognizione situata* o *distribuita* hanno cercato di superare i vecchi modelli quanto a tempi di risposta («azione in tempo reale») ma soprattutto quanto a relazione con l'ambiente. Esempi importanti di questa «nuova I.A.» sono le ricerche di Rodney Brooks[155] e, legato ad esse, l'approccio esistenzialistico nel campo dell'Intelligenza Artificiale. Martin Heidegger, con la sua riflessione sul rapporto tra uomo e natura secondo cui la mente è un prodotto naturale solamente in un'ottica dominata dalla tecnica ma non in una reale, e la teoria di Vygotskij secondo cui «la cognizione umana è distribuita tra gli individui e gli artefatti presenti nell'ambiente esterno»[156], hanno

[149] Cfr. Ivi, pp. 311-330.

[150] Prefazione di G. Tononi, in G. M. Edelman (1995), pp. XXVIII-XXIX.

[151] Ivi, p. XXIX.

[152] Ibidem.

[153] J. Horgan (1998), p. 261.

[154] G. M. Edelman (1993b), p. 294.

[155] Le realizzazioni robotiche di Brooks sono state considerate come l'applicazione di una teoria del mentale proposta recentemente da Riccardo Manzotti e Vincenzo Tagliasco, di cui si parlerà nell'ultimo capitolo.

[156] E. Burattini-R. Cordeschi (2001), p. 263. Si aggiunge nella stessa pagina che «secondo questo approccio, un sistema "cognitivo" è un sistema composto da un individuo che interagisce con un insieme di strumenti oppure da un gruppo di individui che interagiscono tra loro mediante artefatti esterni. Un sistema cognitivo così definito fornisce le risorse informative necessarie all'esecuzione di un'attività. Queste risorse possono es-

ispirato dal lato teorico le critiche al funzionalismo e al materialismo da parte di Herbert Dreyfus e dal lato tecnico le sperimentazioni di Brooks culminanti nella «robotica basata sul comportamento» (*behavior-based robotics*).

Per Dreyfus il comportamento umano è molto più complesso di quello di una macchina giacché quest'ultimo è guidato da un programma che risiede all'interno della macchina stessa mentre per quanto concerne l'uomo le informazioni sono «depositate nell'ambiente esterno (naturale e sociale). Un essere intelligente non può essere semplicemente un programma; deve essere – e qui si avverte un'eco del pensiero fenomenologico di M. Merleau-Ponty – un *corpo* che interagisce con altri corpi»[157]. Con queste premesse Biuso può affermare che «l'Intelligenza Artificiale "classica" viene progressivamente sostituita da prospettive (come la NFAI, *New Fashioned Artificial Intelligence*, e la LAI, *Light Artificial Intelligence*), assai più attente della prima agli aspetti non solo computazionali della mente. E dal convergere della NFAI con le neuroscienze e con alcune delle proposte più solide della filosofia del Novecento, è nata la nuova *embodied cognitive science*, la scienza cognitiva incarnata. Per comprendere la mente, insomma, non si può assolutamente astrarre dal cervello, dal corpo, dal mondo. Ed è per questo che qualunque progetto faustiano di creare l'*homunculus* sotto forma di Intelligenza Artificiale è destinato allo scacco»[158]. Biuso aggiunge efficacemente al riguardo che «la vita, il pensare e l'essere dell'uomo formano un tutto nel quale convivono e si scambiano continue informazioni: a) la struttura chimico-molecolare, b) le leggi fisiche, c) i processi biologici, d) l'identità psichica, e) le relazioni sociali, f) i paradigmi e le visioni del mondo culturali. Isolare uno di questi elementi significa immaginare un'intelligenza artificiale priva di relazioni con il corpo, con la natura e con gli altri, significa produrre qualcosa di assai lontano dalla concreta, quotidiana, operativa,

---

sere rappresentate internamente, nella mente dell'individuo, o nell'ambiente esterno, in quest'ultimo caso possono trovare varie forme di rappresentazione. La distribuzione di queste risorse è la base da cui partire per comprendere il funzionamento dell'interazione e per imparare a progettarla».

[157] S. Nannini (2002), p. 145. Su Dreyfus, cfr. anche E. Carli (2003), pp. 87-107 (qui non Herbert ma Hubert L.) dove lo stesso Dreyfus aggiunge: «sono convinto che, per quanto l'intelligenza artificiale possa svilupparsi e progredire, non sarà mai in grado di elaborare un programma capace di interagire con il mondo e di orientarsi con successo nelle scelte complesse, ma talvolta anche semplici e banali, che caratterizzano la vita quotidiana degli uomini. Credo che, sul piano teoretico, l'intelligenza artificiale si stia rivelando un enorme fallimento» (Ivi, p. 106).

[158] A. G. Biuso (2004), p. 68.

biologica e culturale intelligenza umana. Tale intelligenza non è programmabile in una mente artificiale che rimanga separata dallo spessore profondo e dall'esperienza integrale della corporeità»[159]. In questo senso «una critica seria del determinismo biologico e del funzionalismo computazionale passa, dunque, certamente attraverso la difesa della varietà della vita culturale umana studiata, però, non nella sua separatezza ma nel suo ancoraggio nella vita biologica individuale»[160].

Brooks tenta invece di riformulare i problemi dell'I.A. classica per fornire alla macchina un'intelligenza operativa superiore, in un'ottica «eliminativista». Il sistema intelligente, per Brooks, deve possedere un sistema sensoriale e un sistema motorio, cioè un *corpo*, «deve essere situato in un ambiente esterno e deve poter interagire direttamente con tale ambiente, piuttosto che mediante rappresentazioni simboliche elaborate da uno sperimentatore umano. L'architettura fondamentale di un sistema di questo tipo non deve essere organizzata in una sequenza di moduli funzionali che vengono attivati in successione […] come nella visione classica, ma piuttosto in una serie di moduli comportamentali collegati in modo diretto all'ambiente esterno»[161], moduli come «manipolazione dell'ambiente», «costruzione di mappe», «esplorazione», «evitamento degli ostacoli» e «locomozione». La differenza sostanziale consiste nel fatto che «in questo approccio il comportamento complessivo del sistema non è visto come qualcosa che viene prodotto dal sistema direttamente, ma piuttosto come qualcosa che *emerge* dall'interazione tra le sotto-componenti del sistema e tra il sistema e l'ambiente. Ad esempio, un comportamento che consente al robot di inseguire una fonte luminosa in movimento può emergere dall'interazione tra due moduli comportamentali del sistema che consentono al robot di evitare gli ostacoli e di avvicinarsi a una fonte luminosa»[162].

---

[159] Ivi, p. 31.

[160] V. Gessa Kurotschka (2004), p. 123.

[161] E. Burattini-R. Cordeschi (2001), p. 125.

[162] Ivi, p. 127. Recentemente sono nate le cosiddette «reti neurali miste» che, basandosi sulla scoperta secondo cui «le cellule nervose possono imparare a distinguere tra sequenze di impulsi elettrici, vale a dire memorizzarle» interfacciando i neuroni con materiali artificiali, per esempio silicio, costituiscono «l'ultima frontiera della bionica. […] Questi esperimenti [indicano che] le cellule nervose [hanno] la proprietà di interconnettersi, di interfacciarsi con materiali non biologici […] di estrarre informazioni da ambienti e realtà artificiali: […] i neuroni sono dotati della capacità di associarsi in reti intelligenti e di interfacciarsi con circuiti artificiali pur-

Il funzionalismo deve affrontare le stesse critiche, già discusse in precedenza, che gli antiriduzionisti rivolgono ai materialisti, oltre quella sopra citata di Dreyfus. Ma a queste si aggiungono quelle di Roger Penrose e di John Searle. Sia il matematico che il filosofo condividono la tesi per cui la mente dell'uomo non è una macchina di Turing e la componente essenziale del mentale è la *capacità ermeneutica* o, nei termini da loro usati, la capacità di interpretare semanticamente i simboli esterni. Searle ha puntato inoltre l'attenzione sulla componente biologica dell'umano criticando l'I.A. che ha chiamato «forte» proprio per la scarsa attenzione da questa posta nei confronti del «substrato materiale», dando rilevanza al solo software. Con il famoso *Gedankenexperiment*[163] della «stanza cinese», Searle suggerisce che il computer che elabora simboli non comprende realmente ciò che processa giacché manca l'intenzionalità, mancano i «ruoli causali» che caratterizzano la materia cerebrale. Egli afferma così che «gli stati e gli eventi mentali sono letteralmente un prodotto del funzionamento de cervello, mentre il programma non è un prodotto del calcolatore»[164].
Penrose, appoggiando la tesi di Searle,[165] tenta di porre la differenza tra uomo e computer in termini logico-matematici.[166] Partendo dal presupposto, corretto, che l'algoritmo – base

---

ché questi siano in grado di recepire i loro segnali e ritrasmetterli alla rete neurale» (A. Oliverio 2004, pp. 138-139).

[163] Cfr. D. R. Hofstadter-D. C. Dennett (2001), pp. 341-360 e T. O'Connor-D. Robb (2003), pp. 332-352, entrambi contenenti l'articolo di Searle «Menti, cervelli e programmi» (*Minds, brains and programs*).

[164] D. R. Hofstadter-D. C. Dennett (2001), p. 357.

[165] Cfr. R. Penrose (1997), p. 27 e pp. 39-47.

[166] Un tentativo analogo è stato tentato da J. R. Lucas nel 1961: secondo la sua tesi, una mente cosciente può evitare le implicazioni del teorema di Gödel perché può «riflettere su se stessa e criticare le proprie operazioni, e a questo scopo non ha bisogno di alcuna parte in più [mentre] una macchina può essere costruita, per così dire, in modo da "prendere in considerazione" le proprie operazioni, ma non può farlo senza diventare con questo una macchina diversa, cioè la vecchia macchina con in più una "parte nuova"» (D. R. Hofstadter 1984, p. 422). Dunque per Lucas «qualunque sistema che non sia sconfitto dal problema di Gödel *eo ipso* non è una macchina di Turing, cioè non è una macchina nel senso legittimo del termine» (Ivi, p. 423) chiarendo che «noi siamo sempre *fuori* dal sistema, e che da là fuori possiamo sempre compiere l'operazione della "gödellizzazione", la quale fornisce qualcosa di cui il programma, dall'interno, non può vedere la verità» (Ivi, p. 510).
Douglas Hofstadter risponde all'argomento di Lucas nel 1979 dicendo a p. 513 che «in realtà, proprio il fatto che non riusciamo a scrivere un programma per la "gödellizzazione" deve farci sospettare che neppure noi sapremmo eseguirla in tutti i casi. Un conto è dimostrare in astratto che la gödellizzazione "si può fare"; e un altro conto è saperla fare in ciascun caso particolare. Di fatto, a mano a mano che cresce la complessità dei sistemi formali (o dei programmi), la nostra capacità di "gödellizzazione" s'indebolisce e prima o poi comincia

di ogni computazione elettronica – ha una componente mentale umana che dunque lo fa dipendere dall'operatore umano, Penrose avanza l'ipotesi secondo cui ciò che è algoritmico nella mente umana è la parte non cosciente, mentre ciò per cui l'uomo adopera la coscienza (il «senso comune», il «giudizio di verità», la «comprensione» e l'«apprezzamento artistico») non è riducibile a processi di tipo algoritmico: «la *formazione del giudizio*, che sto sostenendo essere il contrassegno della coscienza, è *di per sé* qualcosa che gli studiosi dell'IA non avrebbero alcuna idea di come programmare al computer»[167]. Il matematico propone inoltre una connessione tra il mentale e la fisica quantistica, affermando cautamente: «si potrebbe […] congetturare che da qualche parte, nella profondità del cervello, si trovino cellule sensibili a singoli quanti. Se questa congettura dovesse venire confermata dall'osservazione, la meccanica quantistica si troverebbe a essere implicata in modo significativo nell'attività cerebrale»[168].

Oggetto di indagine solo teorica ma fisicamente possibile delle applicazioni delle teorie simili a quelle di Penrose è il «computer quantistico» che invece di elaborare l'informazione

---

a vacillare. Ciò deve accadere poiché […] *non* possediamo alcun algoritmo che ci spieghi come eseguirla. Se non riusciamo a dire *esplicitamente* che cosa comporta applicare il metodo di Gödel in tutti i casi, allora prima o poi ciascuno di noi incontrerà un caso talmente complicato che semplicemente non riuscirà a capire da che parte cominciare». Ammette che «certamente un programma ha la possibilità di modificare se stesso, ma questa modificabilità dev'essere inerente al programma fin dal principio e quindi non può essere considerata un esempio dell'"uscire dal sistema" […]: sfuggire alle regole è per [il programma] altrettanto impossibile quanto lo è per un essere umano infrangere con la volontà le leggi della fisica» (Ivi, pp. 515-516). Hofstadter comunque propone di «incorporare in un programma di calcolatore una capacità parziale di "uscire da se stesso"» (Ivi, p. 516) come avviene, continua nella stessa pagina Hofstadter, «da un sottosistema del proprio cervello a un altro sottosistema più vasto [creando la sensazione] di uscire da se stessi». Ma così non eviterebbe la sottile difficoltà messa in luce da Lucas e da tutti coloro che pongono il teorema di Gödel come linea di demarcazione tra macchina e macchina pensante, nonché tra computer e uomo.

[167] R. Penrose (1997), p. 520.

[168] Ivi, p. 506. Ma continua nella stessa pagina: «pare che, nella migliore della ipotesi, tutto ciò che potremo ottenere dalla meccanica quantistica sarà un'incertezza circa la questione se un neurone scaricherà o no, ed è difficile vedere in che modo questo fatto potrebbe risultare di grande utilità per noi». E a p. 290 dice che «il fenomeno stesso della coscienza potrebbe essere incomprensibile in termini puramente classici», di fisica classica. Propone in merito il principio detto «antropico» per cui «la natura dell'universo […] è fortemente vincolata dalla richiesta che esseri intelligenti come noi stessi debbano essere realmente presenti per osservarlo» (Ivi, p. 513). Sulla fisica quantistica è utile H. Putnam, *Osservazioni di un filosofo sulla meccanica quantistica*, in H. Putnam (1993a), pp. 150-177.

in termini binari («0» o «1») è in grado di processare valori intermedi (il «qubit»). I computer a base «analogica» svolgono un compito simile ma i computer quantistici, oltre che per le prestazioni in termini di operazioni al secondo, differiscono da questi nel fornire un risultato non analogico ma intermedio: «la procedura di un computer quantistico è quindi analogica ma il suo risultato è digitale e ciò sommerebbe la precisione del digitale con la flessibilità dell'analogico»[169].

Nonostante la «nuova» I.A. abbia notevolmente migliorato i progetti precedenti, le reti artificiali di nuova generazione presentano ancora gravi limiti (trascurando le più profonde difficoltà di ordine filosofico). Burattini e Cordeschi analizzano le principali difficoltà tecniche indicando nuove soluzioni per ognuna di esse.[170] Eppure, come ricorda Hofstadter, sembra che sui progressi nel campo dell'I.A. valga un «teorema di Tesler» per cui «l'IA è tutto ciò che non è ancora stato fatto»[171], nel senso che «non appena si riesce a programmare qualche funzione mentale, immediatamente si smette di considerarla un ingradiente essenziale del "vero pensiero". Il nucleo ineluttabile dell'intelligenza è sempre in quell'altra cosa che non si è ancora riusciti a programmare»[172]. Forse è così proprio perché la mente non è una *somma* di funzioni le quali, se messe insieme l'una accanto all'altra, *danno come risultato* la coscienza e l'intelligenza, i grandi obiettivi dell'Intelligenza Artificiale. Forse solo cambiando gli orizzonti limitanti e limitati dei *ricercatori informatici di intelligenza* è possibile compiere passi concreti verso la comprensione della natura della mente. Ma è necessario, forse, un cambio *totale* di prospettiva.

Lasciando spazio nel prossimo capitolo a questo argomento, completiamo questo approfondendo, nei limiti consoni, il rapporto tra la filosofia della mente e le neuroscienze.

---

[169] A. G. Biuso (2004), pp. 54-55.

[170] E. Burattini-R. Cordeschi (2001), pp. 112-113.

[171] D. R. Hofstadter (1984), p. 649.

[172] Ibidem. Utile a proposito è l'elenco delle attività dell'I.A. proposto da Hofstadter. Esso comprende traguardi nei seguenti campi fondamentali: traduzione automatica, "comprensione" e produzione di linguaggio naturale (si tratta di risposte a domande in specifici ambiti, analisi di frasi complesse, costruzione di parole e frasi, etc.), dimostrazione di teoremi in vari campi della matematica (logica simbolica, geometria elementare) e manipolazione simbolica di espressioni matematiche, "visione" di scritture e di scene generiche (localizzazione in fotografie di oggetti specificati in precedenza, lettura di testi), "comprensione acustica" di segmenti di fonemi, "creazione di pensieri originali e di opere d'arte" (si tratta di "arte col calcolatore"), "pensiero analogico" (costruzione di dimostrazioni di un dominio della matematica basate su quelle di un dominio affine) e altri (Ivi, pp. 649-651).

La teoria edelmaniana di cui sopra si è discusso rappresenta il ponte di contatto esemplare tra temi cari alla filosofia della mente e quelli che costituiscono materia di scienza, in questo caso la neurobiologia. E se Gerald Edelman ha contribuito notevolmente in campi filosofici e ingegneristici, anche gli studi di Antonio Damasio rivestono sempre più importanza nelle ricerche dell'I.A. in quanto è ormai consolidata l'importanza delle emozioni sia nelle scelte umane considerate razionali sia in un prodotto artificiale che voglia simulare al meglio l'intelligenza umana. Ecco a proposito il resoconto di Damasio datato 1994: «per il tema della "coscienza" […] sono convinto che la neurobiologia possa cominciare ad accostarvisi. Alcuni filosofi (tra i quali John Searle, Patricia Churchland e Paul Churchland) hanno esortato i neurobiologi a studiare la coscienza, e sia filosofi sia neurobiologi (per esempio Francis Crick, Daniel Dennett, Gerald Edelman, Rodolfo Llinás) hanno cominciato a formulare teorie»[173]. *L'errore di Cartesio* di cui parla Damasio è il dualismo radicale di corpo e mente. Non solo è necessario reintegrare la funzione del corpo (come l'approccio esistenzialista e fenomenologico propone) nei processi mentali ma va rivisto il concetto di mente legata al solo cervello. Il neurobiologo ha mostrato nelle sue ricerche che «i circuiti neurali rappresentano con continuità l'organismo, mentre esso è perturbato da stimoli provenienti dall'ambiente fisico e da quello socioculturale, e mentre agisce su tali ambienti. Se l'oggetto fondamentale di tali rappresentazioni non fosse un organismo ancorato nel corpo, potremmo sì avere qualche forma di mente, ma dubito che sarebbe la mente che abbiamo. […] il contributo del corpo al cervello non si riduce agli effetti modulatori o al sostegno delle operazioni vitali, ma comprende anche un *contenuto* che è parte integrante del funzionamento della mente normale»[174].

[173] A. R. Damasio (1995), pp. 320-321.

[174] Ivi, pp. 308-309. Aggiunge a p. 310: «perché lo stato biologico del sé si verifichi, bisogna che siano integri e in piena attività numerosi sistemi cerebrali e anche numerosi sistemi corporei. Supponete che vengano recisi *tutti* i nervi che portano al corpo i segnali del cervello; il vostro stato corporeo cambierebbe in modo radicale, e così pure, di conseguenza, la vostra mente. Anche un arresto parziale degli scambi cervello-corpo (come si ha in pazienti con lesioni al midollo spinale) modifica lo stato della mente». Non è un caso che subito dopo parli del suggestivo esperimento mentale di Putnam del «cervello in vasca» criticandone gli aspetti collaterali: il cervello senza corpo modifica la mente. Damasio suggerisce inoltre che «la soggettività emerga […] quando il cervello produce non solo immagini di un oggetto, non solo immagini delle risposte dell'organismo all'oggetto, ma immagini di un terzo tipo, cioè immagini di un organismo nell'atto di percepire un oggetto e di rispondervi» (Ivi, p. 329). Studi simili a quelli neurobiologici di Damasio sono stati compiuti anche dallo psicologo Daniel Goleman approfondendo proprio il legame tra emozione e intelligenza (cfr. D. Goleman 1999).

Il biologo Steven Rose invece ha puntato l'attenzione sulla memoria: criticando l'analogia mente-software («i cervelli […] non lavorano sull'informazione nel senso dei computer, ma sul *significato*»[175]), egli afferma che gli apparati sensoriali sono capaci di «un'*auto-organizzazione dipendente dall'attività* [per cui] cellule cerebrali che modificano le loro proprietà in questo modo adattivo in risposta a contingenze ambientali presentano chiaramente una proprietà piuttosto vicina alla memoria»[176] definita come «qualcosa di cui dobbiamo inferire l'esistenza come processo intermedio per connettere l'apprendimento al richiamo. […] In altri termini, all'interno dell'organismo dev'essere avvenuto un qualche tipo di registrazione, in virtù del quale l'informazione acquisita durante l'apprendimento è stata memorizzata in forma tale da essere resa disponibile per modificare il comportamento successivo»[177].

Di Francesco nota che l'apporto della ricerca scientifica alla riflessione sul mentale ha condotto da un lato alla scienza cognitiva che «distribuisce i compiti cognitivi anche al di là dei limiti fisici dell'agente [mentre,] dall'altro, lo straordinario sviluppo della neuroscienza […] tende a aumentare sempre di più le conoscenze che possediamo sul cervello e sul modo in cui esso supporta la mente, riducendo sempre di più il pensiero alla sua base materiale»[178] ed un esempio di questo genere è il lavoro di Jean-Pierre Changeux. Nel suo interessante *L'uomo neuronale*, dopo aver illustrato la storia dello studio del cervello in filosofia e nella scienza dall'antico Egitto ad oggi (il testo è del 1983), Changeux espone il suo materialismo democriteo: «l'attività nervosa, evocata o spontanea, e la sua propagazione nelle reti di neuroni si spiegano in fin dei conti con proprietà *atomiche*. […] Il sistema nervoso si compone – e impiega per funzionare – della stessa "materia" del mondo inanimato»[179], proponendo di sostituire gli «atomi psichici» con «macromolecole responsabili della comunicazione nevosa»[180]. Anche se, con Edelman, Damasio e Rose, rifiuta l'analogia funzionalista mente-software, afferma: «Bergson, in *Matière et Mémoire* [(*Materia e memoria*)], scriveva che "il sistema nervoso non ha nulla di un apparato che servirebbe a fabbricare o anche a preparare rappresentazioni". La tesi sviluppata [da Changeux] è l'esatto contrario di

[175] S. Rose (1994), p. 117.

[176] Ivi, p. 178.

[177] Ivi, p. 170.

[178] M. Di Francesco (1996), p. 193.

[179] J. P. Changeux, p. 116.

[180] Ivi, p. 117.

quella bergsoniana [giacché] l'encefalo dell'uomo, di cui si sa che *contiene*, nell'organizzazione anatomica della sua corteccia, rappresentazioni del mondo circostante, è anche *capace* di costruirne e di utilizzarle nei propri calcoli»[181].

Nonostante la considerazione funzionale delle strutture biologiche (con le difficoltà che comporta la definizione di «funzione» in biologia) non può essere ricondotta alla nozione riduzionistica per cui la struttura biologica è una macchina, il riduzionismo risulta carente, «riduttivo»: le funzioni dell'organismo – a livello cellulare – si modificano in base alla relazione con l'esterno. Le parti funzionali della macchina sono statiche. L'essere biologico non è riducibile alla macchina (al contrario di quanto pensa Descartes) perché esso si modifica con livelli di tolleranza incredibilmente maggiori rispetto alle macchine, grazie soprattutto alla complessità strutturale. L'*omeostaticità*, l'adattamento ambientale, è la caratteristica tipica degli organismi viventi e li differenzia dalle macchine.[182]

Come dice Paolo Legrenzi, «anima, mente e cervello sono tre livelli di descrizione separati. La scienza dell'ultimo secolo ha eliminato il primo termine della triade, ma ha dimostrato la fecondità di studiare le interazione tra gli altri due. Le frontiere della ricerca contemporanea si spostano sempre più avanti mettendo in rapporto lo studio della mente con quello sul funzionamento del cervello»[183]. Un caso esemplare in cui la filosofia e la ricerca tecnologica e neuroscientifica si sostengono e confermano vicendevolmente è quello del connessionismo e il *neo-eliminativismo*. I coniugi Churchland, che hanno parlato di "neurofilosofia", hanno contribuito allo sviluppo teorico del «connessionismo californiano in alternativa all'ortodossia del MIT basata su Chomsky, Fodor, l'intelligenza artificiale classica e, più in generale, l'analogia mente-computer»[184], nella misura in cui le ricerche sulle reti neurali hanno contribuito al completamento della loro posizione filosofica. Infatti la loro teoria della mente propone un paradigma alternativo a quello linguistico, per cui il pensiero è fondamentalmente un linguaggio (in particolare per Fodor), ovvero sostiene che le attività intellettuali insieme a tutte le altre svolte dal cervello umano e animale siano riconducibili al coordinamento senso-motorio, quella capacità di trasformare un input sensoriale in uno motorio. E' qui che viene in soccorso la ricerca della nuova I.A.: «le reti neurali connessionisti-

[181] Ivi, p. 152.

[182] Considerazioni suggerite dal Prof. G. Camardi.

[183] P. Legrenzi (2002), p. 37.

[184] S. Nannini (2002), p. 183.

che mostrano che il coordinamento senso-motorio potrebbe essere ottenuto dal cervello, in modo biologicamente plausibile, non mediante un calcolo matematico o, più in generale, un 'meccanismo inferenziale' basato sulla *manipolazione di simboli* grazie a regole sintattiche, bensì attraverso una *trasformazione di vettori*»[185]. Nel tentativo di spiegare le attività intellettive e sensoriali del cervello umano, naturalmente integrando le ultime scoperte in campo neuroscientifico, i Churchland assegnano il compito di trasformare un input sensoriale in un output sensoriale (che avviene a livello neuronale in ogni organismo vivente) alle reti connessionistiche. Va precisato che «il comportamento di una rete nella sua interezza è determinato dalla configurazione della connettività esibita dalle sue unità in un dato momento e dallo stato globale di attivazione della rete in quel momento [e questo stato è a sua volta indicato da un vettore (lista ordinata) di numeri reali]. Tutti i possibili stati globali di attivazione della rete sono rappresentati come il suo spazio di attivazione, la cui dimensionalità è esattamente equivalente al numero di unità della rete. […] Una serie temporale di stati globali di attivazione traccerà una traiettoria (*path*) attraverso uno spazio-vettore. Il processo dell'informazione nella rete è rappresentato come l'evoluzione nel tempo di configurazioni (*patterns*) globali di attivazione»[186]. A questo punto «se costruiamo […] un secondo spazio degli stati dove [rappresentiamo] i *pattern* motori, allora il problema del coordinamento senso-motorio si riduce a quello di un calcolo vettoriale [che metta in corrispondenza biunivoca] ogni punto del primo spazio con un punto del secondo»[187]. Il calcolo, così diventato matriciale, è matematicamente trattabile ma concretamente troppo lungo da effettuare per un uomo. Le reti artificiali, secondo i Churchland, possono svolgere questo calcolo in modo automatico, affermando così che l'unica differenza tra l'uomo (o l'animale in genere) e la macchina connessionistica risiede nella *quantità* di unità presenti nella rete, non nella complessità strutturale della stessa.

---

[185] Ivi, p. 192.

[186] B. P. McLaughlin, *Computationalism, connectionism, and the philosophy of mind*, in L. Floridi (2004), p. 143. Nel testo: «The behavior of a network as a whole is a consequence of the pattern of connectivity exhibited by its units at a time and the global activation state of the network at that time. […] The set of all possible global activation states of a network is its activation space, whose dimensionality is exactly equal to the number of units in the network. […] A temporal series of global activation states will trace a path through a vector space. Network information processing is characterized as the evolution through time of global patterns of activation».

[187] S. Nannini (2002), p. 193.

Le discusse teorie dei Churchland pongono con insistenza il naturalismo scientifico alla base delle teorie filosofiche riguardanti il mentale, ma è pur vero che «le ricerche più avanzate sulla mente non si pongono oggi lo scopo di costruirne il modello computazionale complessivo. La complessità della mente, soprattutto se la si assume in tutti i suoi aspetti, anche quelli fenomenici, i così detti *qualia*, non è realisticamente riproducibile attraverso un modello computazionale»[188]. Difficoltà sorgono anche per il cognitivismo: Patrizia Tabossi, trattando del rapporto tra neuroscienze e cognitivismo, nota che «la difficoltà di considerare le neuroscienze come parte integrante della scienza cognitiva non dipende tanto […] dall'accento posto da questa sugli aspetti funzionali delle attività intelligenti, quanto piuttosto dalla sfasatura spesso esistente fra il livello di astrazione delle teorie proposte dalle varie scienze cognitive e i modelli delle neuroscienze. In linea di principio, queste ultime dovrebbero fornire criteri che consentano di restringere le possibilità teoriche dei modelli cognitivi, stabilendo i vincoli biologici dei diversi organismi. Purtroppo però, mentre le neuroscienze hanno raggiunto livelli di conoscenza molto raffinati su alcuni meccanismi neuronali, in molti casi la loro capacità di spiegare come tali meccanismi siano integrati in processi cognitivi complessi è ancora limitata»[189].

La mente (e con essa la coscienza, l'intenzionalità e l'intelligenza) dunque non si lascia ridurre a qualcosa di trattabile esaustivamente a livello scientifico, nonostante la scienza stia tentando di coglierne il segreto. Le vie sembrano essere due: o la mente nasconde dei segreti che non permettono, con l'attuale livello tecnologico e culturale umano, di svelarne gli intrighi (ma il segreto del mentale è lì che aspetta di essere scoperto) *oppure* è il concetto stesso di «mente» che, sfuggendoci a causa delle fuorvianti categorie di interpretazione che utilizziamo oggi nei suoi confronti, non sarà mai colto. Forse perché *la realtà ultima della mente è la realtà ultima*, ontologicamente intesa.

[188] V. Gessa Kurotschka (2004), p. 120.

[189] P. Tabossi (1988), pp. 159-160.

«I confini dell'anima, per quanto lontano tu vada,
non li scoprirai, neanche se percorri tutte le vie:
così abissalmente si dispiega»
Eraclito[1]

Capitolo 3.

Della realtà ultima

Come ha detto Kant, «già dai tempi più antichi della filosofia i filosofi speculativi hanno posto, oltre gli esseri sensibili o fenomeni, costituenti il mondo sensibile, certi particolari esseri intelligibili (*noumeni*) che dovrebbero costituire un mondo intelligibile e, poiché essi facevano una cosa sola della parvenza (o fenomeno) e dell'apparenza illusoria (ciò che può perdonarsi ad un'età non ancora raffinata), a questi soli esseri intelligibili attribuivano una vera realtà»[2]. La realtà ultima è stata spesso ricercata al di fuori del visibile; ma lo è stata necessariamente agli albori della filosofia. Eraclito stesso, dicendo che «l'Origine ama nascondersi»[3], distingueva la realtà dei sensi da quella «ultima»[4]. La posizione ontologica eraclitea si può oggi definire un *monismo neutrale* in quanto ritiene unica la "sostanza" ma duplice o molteplice il modo in cui la sostanza può presentarsi. Prima di Eraclito, sin da Talete si indagava la "realtà ultima" delle cose, chiamata «αρχή» da Anassimandro. Con Parmenide si ha, per la prima volta, la concezione che tutto sia un'unità;[5] l'Essere parmenideo non è «principio» perché non esiste «principiato» in quanto la natura è unica ma è la «doxa» a ritenerla molteplice:[6] l'eleatismo (Parmenide e i suoi seguaci, Zenone e Melisso) non

[1] Eraclito (2005), p. 181.

[2] I. Kant (1995), p. 147.

[3] Eraclito (2005), p. 191.

[4] «Ogni sostanza, per Eraclito, "è nello stesso tempo un'unica sostanza e due sostanze opposte" (Diano-Serra FT 144), a seconda dell'interazione che si attua di volta in volta. […] qui si afferma […] una metafisica della cangianza radicale» (Ivi, p. 147).

[5] Già in Eraclito era presente in realtà questa concezione: «è saggio ammettere che tutto è uno» (G. Reale 2004, p. 117) ma ancora in forma non chiara quanto quella parmenidea.

[6] Cfr. Ivi, p. 182.

concepiva diversi piani ontologici dell'Essere.[7] Baruch Spinoza sembra condividere questo assunto aleatico, quando afferma che l'unica sostanza esistente è Dio, mentre *pensiero* ed *estensione* sono due dei suoi infiniti «attributi».[8] Dunque passando da Eraclito e Parmenide fino a Spinoza, il monismo neutrale è stato considerato un modo alternativo di cogliere la realtà sia al materialismo sia all'idealismo. Una concezione più moderna di questa posizione filosofica si ha con David Hume: con lui per la prima volta appare la nozione di «entità neutrali», entità ultime che non possono essere descritte né come mentali né come materiali e soprattutto l'idea che il corpo e la mente sono riducibili a (e formati da) queste entità.[9]
Questa concezione humeiana giungerà a Ernst Mach, William James e Bertrand Russell, i maggiori esponenti del monismo neutrale, con intenti inizialmente riduzionistici e in alcuni casi eliminativistici.[10] Il primo (indipendentemente ma contemporaneamente a Richard Avenarius) rintraccia l'entità neutrale nell'«esperienza pura»[11], per il secondo è il «flusso di coscienza» (continuo delle esperienze) a presentare entità materiali e mentali come distinte (per il loro modo di essere raggruppate) non essendole in natura,[12] il terzo definisce le entità

---

[7] Parmenide ammise che la «*doxa* si *deve* costruire su una *polarità* di principi, che tuttavia devono essere ricondotti all'unità dell'Essere, per non introdurre il Nulla come uno dei principi del fenomenico» (M. Andolfo 1999, p. 57). Poi i naturalisti (Empedocle e Anassagora) tenderanno al monismo materialistico ma, a differenza degli atomisti, ignorando ancora la distinzione categoriale di materia e spirito.

[8] I corpi sono «modi» dell'estensione divina, le menti e le loro idee sono modi del pensiero divino, correlati da un parallelismo logico: mente e corpo sono identici per virtù divina, senza interazione causale.

[9] Cfr. L. Stubenberg (2005), cap. 2. Riguardo Kant, la sua posizione in merito è particolare: è dualista nel senso che distingue corpo e mente come sostanze diverse, ma ammette che entrambe sono manifestazioni della «cosa in sé» (l'oggetto trascendentale) la cui natura è ignota. Può darsi che "in quanto cosa in sé" non ci sia differenza tra corpo e mente: può essere considerata una forma di «monismo neutrale trascendentale».

[10] Cfr. Ivi, § 5.1.

[11] Come per Kant, in *L'analisi delle sensazioni e il rapporto tra fisico e psichico* (1900), Mach afferma che il limite conoscitivo è l'esperienza che ha elementi né fisici né mentali ma o gli uni o gli altri in base al modo in cui sono raggruppati o considerati: la materia conosciuta come tale è una connessione regolare di elementi (sensazioni). L'Empiriocriticismo di Mach proviene da quello di Avenarius per cui l'esperienza pura viene prima di ogni distinzione tra fisico e psichico. In fisica Mach presenta una forma di fenomenismo berkeleyano per cui «ciò che rappresentiamo a noi stessi dietro le apparenze esiste *solamente* nel nostro intelletto» (J. Losee 2001, p. 158). Su Mach e Avenarius, cfr. G. Reale-D. Antiseri (1998), pp. 305-316.

[12] Afferma James nel 1912: «mente e materia, soggetto e oggetto, pensiero e cosa, rappresentazione e rappresentato sono dunque interpretati come risultato di differenti raggruppamenti di pura esperienza» (L. Stuben-

neutrali come aveva fatto Hume «essenza neutrale» (*neutral stuff*).[13]

Russell rappresenta la fonte contemporanea fondamentale di questa posizione filosofica. In varie opere ma in particolare in *L'analisi della mente* del 1921, Russell espone la sua «conversione» al monismo neutrale, anche se in *Analysis of Matter* lo abbandona per le influenze di Feigl e Maxwell.[14] Nelle quindici lezioni che compongono il testo del 1921, Russell vuole confutare la teoria per cui con «mentale» si intende un «qualcosa di peculiare detto "coscienza" e concepito o come una relazione con gli oggetti, o come una qualità intrinseca dei fenomeni psichici»[15]. I termini «materia» e «mente» sono imprecisati sia in filosofia che nella scienza: «la sostanza di cui è composto il mondo della nostra esperienza non è, a mio avviso, né mente né materia, ma qualcosa di più primitivo delle due […] come un antenato comune»[16]. Identificando i realisti con i "soppressori del contenuto" e gli idealisti con "soppressori dell'oggetto", riconosce il contributo di James alla propria posizione: questi aveva respinto il termine «coscienza» come entità affermando invece che «c'è una funzione nell'esperienza che i pensieri compiono e per il cui compimento questa qualità di essere è

---

berg 2005, cap. 2. Nel testo: «mind and matter, knower and known, thought and thing, representation and represented are then interpreted as resulting from different groupings of pure experience»).

[13] Oltre a questi tre autori fondamentali, hanno aderito al monismo neutrale Joseph Petzoldt (seguace di Mach e Avenarius), Ralph Barton Perry (tra i «nuovi realisti americani», seguì il pensiero di James), Edwin B. Holt (anche lui tra i realisti, ispirò Russell coniando il termine «neutral stuff»), H. H. Price (posteriore a Russell, per Price erano neutrali i dati sensoriali) e, infine, T. P. Nunn (che influenzò Russell per l'*extramentalità* della percezione). Cfr. L. Stubenberg (2005). Due autori fondamentali più noti dei precedenti sono Theodor Fechner e Wilhelm M. Wundt. La loro posizione è meglio definita in Stubenberg come «teoria del doppio aspetto» che, si ammette, è difficilmente distinguibile dal monismo neutrale. Per il primo la realtà può apparire fisica o mentale in base al punto di osservazione, per il secondo dietro l'apparenza sensibile si cela la *volontà* schopenhaueriana. A parte va citato Husserl: il suo può definirsi un monismo neutrale «empirico» per cui nella coscienza empirica i dati dell'esperienza hanno un'unica natura, ma in base alle «essenze» che esemplificano, essi vengono successivamente distinti come materiali o spirituali. Nannini precisa al riguardo che la realtà empirica è connotata per Husserl da questa natura unica e al tempo stesso duplice (*Körper* e *Leib*, mondo della vita e parte psichica) ma essa è solo il «correlato» del soggetto trascendentale (fuori del reale, che in qualche modo il soggetto crea ma distingue da sé); non va perciò dimenticato l'antinaturalismo husserliano lasciando da parte ogni interpretazione monistico-neutrale.

[14] Cfr. L. Stubenberg (2005), cap. 7.

[15] B. Russell (1970), p. 17.

[16] Ivi, p. 18.

invocata. Questa funzione è il conoscere»[17]. Aggiunge inoltre che la coscienza non è l'essenza della mente, ma solo un termine «nel complesso banale e insignificante delle abitudini linguistiche»[18]. Tramite la teoria della «Prospettiva»[19], separa la fisica dalla psicologia: quando i «particolari» compongono un solo oggetto, tanto da costituire una «singola unità causale» trattabile separatamente, essi sono indagati dalla fisica; se si tratta non di sistemi di particolari ma di «particolari reali», se ne occupa la psicologia.[20] A confermare tutto ciò, Russell specifica il suo «dualismo causale» per cui si dice "fisico" ciò che segue leggi causali fisiche e "mentale" ciò che segue leggi causali mentali; eppure, non essendo precisa la distinzione tra psichico e fisico, Russell stesso ammette che è ancora impossibile conoscere «la» causa di un determinato evento, nonostante la sua categorizzazione sommaria di «sensazioni» e «movimenti volontari».[21] A connotare le sensazioni come qualcosa di «non solo fisico» è la loro caratteristica di «elementi non mnemonici di una percezione»[22];

---

[17] Ivi, p. 26. Con James afferma che contenuto e coscienza (materia e spirito) sono "tipi particolari di relazioni tra parti di pura esperienza". Il termine «esperienza» diventa «neutralità» con i realisti Perry e Holt (di cui sopra) e a partire da questi Russell ritiene le sensazioni siano composte da una sostanza neutra né mentale né fisica, ma ritiene le immagini assolutamente mentali; invece considera assolutamente materiali quei fatti «che non facciano parte di alcuna "esperienza"» (Ivi, p. 28). Ammette così due tipi di leggi causali, fisiche e mentali: le sensazioni sono soggette a entrambe (sono «neutrali») mentre esistono entità «puramente» materiali o mentali in base alle leggi causali (rispettivamente fisiche o mentali) che seguono. Precisa però che la «rigida esattezza universale della fisica» è ancora da provarsi empiricamente e che quindi bisogna mantenere sospeso il giudizio in merito al fatto che la materia vivente segua leggi causali che la materia inanimata invece non segue: Russell concede un'obiezione al determinismo.

[18] Ivi, p. 38.

[19] Ivi, a p. 214, precisa come «i particolari che costituiscono il mondo fisico possono essere raccolti in insiemi in due modi: *a*, un insieme composto di tutti quei particolari che sono aspetti di una data cosa vista da punti diversi; *b*, un insieme composto di tutti quei particolari che sono aspetti di cose diverse, viste da un dato punto. L'insieme di quest'ultimo tipo, considerato in un dato momento, lo chiamiamo "prospettiva"; considerato attraverso un certo periodo di tempo lo chiamiamo "biografia"».

[20] Cfr. Ivi, p. 82. Precisa più avanti che «gli elementi costitutivi ultimi della materia non sono atomi o elettroni, ma sensazioni e altre cose simili alle sensazioni, per quanto riguarda l'estensione e la durata» (Ivi, p. 93). Questi «particolari» sono suscettibili di analisi, solo che per ora non si sa come analizzarli.

[21] Per Russell, le «sensazioni» hanno cause fisiche ed effetti mentali, mentre i «movimenti volontari» hanno cause mentali ed effetti fisici.

[22] Ivi, p. 105.

sono «i punti di intersezione tra spirito e materia»[23], come lo erano per Mach, James e Dewey. Ma la mente non è composta di sole sensazioni bensì anche di "immagini" ovvero una sottoclasse di sensazioni seppure distinte da queste in virtù delle cause e, in misura minore, degli effetti (molto vicino alla distinzione humeiana di "impressioni" e "idee"). Dunque definisce la coscienza come relazione tra un'immagine (o una parola) ed un oggetto, relazione che chiama «significato»[24]. In conclusione, per Russell la «vera metafisica» considera tanto la materia quanto lo spirito formate da una sostanza neutra, «le cui leggi causali non hanno la dualità della psicologia, ma formano la base sulla quale poggiano la fisica e la psicologia»[25].

Più recentemente alcuni autori hanno presentato posizioni molto affini al monismo neutrale; alcuni con decisione, come Chalmers, altri con titubanze, come Putnam, Davidson e Nagel. Il primo condivide con Kenneth Sayre la concezione per cui la realtà ultima sia costituita dall'*informazione*.[26] Chalmers infatti, assumendo che il cervello sia guida del comportamento senza che la mente interferisca, dovette difendersi dall'accusa di epifenomenismo convergendo le sue ipotesi sul mentale nel monismo neutrale,[27] proprio affine a quelle

---

[23] Ivi, p. 108.

[24] Ivi, p. 209. Segue il ragionamento russelliano sul «dato»: non esiste "dato" indipendente dalla credenza (come nella percezione non c'è "dato" finché non c'è "giudizio" di percezione), «una sensazione che va e viene semplicemente non è un "dato"; essa diventa "dato" solo quando è ricordata» (Ivi, p. 215).

[25] Ivi, p. 208. Eppure, nonostante per il filosofo la mente e la materia siano entrambe «costruzioni logiche» (Ivi, p. 221) le cui relazioni sono studiate in parte dalla psicologia in parte dalla fisica, va notato che «tutti i nostri dati, tanto quelli della fisica, quanto quelli della psicologia, sono soggetti a leggi causali psicologiche; ma le leggi causali fisiche, almeno nella fisica tradizionale, si possono formulare solo nei termini della materia, che costituisce insieme una deduzione e una costruzione, ma non è mai un dato. Sotto questo aspetto, la psicologia è più vicina alla realtà della fisica» (Ivi, p. 222).

[26] Cfr. L. Stubenberg (2005), cap. 2. Sayre tenta di concepire come entità neutrali basilari le strutture della scienze matematiche.

[27] Cfr. S. Nannini (2002), p. 155. Chalmers parte dal presupposto che la mente faccia parte della natura, sposando un dualismo delle proprietà per cui servono leggi «psicofisiche» per spiegare la coscienza. Horgan aggiunge la rilevanza del debito, riconosciuto da Chalmers, nei confronti della teoria di Wheeler dell'«*it* from *bit*» (cfr. J. Horgan 1998, p. 280) per cui ogni particella, campo di forza o il continuum spazio-tempo (ogni *it*) «deriva completamente la sua funzione, il suo significato, la sua stessa esistenza – anche se, in alcuni contesti, in maniera indiretta – dalle risposte ricavate dall'apparato sperimentale a domande del tipo sì o no, cioè da scelte binarie, da *bit*» (Ivi, p. 134, dove si riprende lo stesso Wheeler). Per entrambi «l'informazione è una proprietà del reale altrettanto essenziale quanto la materia e l'energia» (Ivi, p. 279).

di Sayre. Quello di Chalmers è un caso esemplare che testimonia come il monismo neutrale rappresenti la via inevitabile per chi non vede la scienza come completa pur non intendendo perdere i contatti con il materialismo: tra antiriduzionismo e materialismo. Hilary Putnam, nel 1973, espone dei dubbi sulla sostanza ultima:[28] ritiene questa fondamentalmente unica e non riducibile né al fisico né al mentale; la «stoffa» di cui parla, che costituisce l'universo, non conduce difatti né al materialismo, dato che l'autonomia del mentale è salvaguardata da ogni legge fisica (che non può spiegare il comportamento umano se non ricorrendo ad un linguaggio psicologico), né al dualismo o all'idealismo, in quanto la struttura mentale è «funzionale» al corpo anche se non ad esso riducibile, come in Aristotele.[29] Per quanto concerne Davidson, Nannini individua un leggero cambio di prospettiva dal monismo anomalo a quello neutrale nello scritto del 1992.[30] Thomas Nagel, già incontrato tra gli antiriduzionisti, è un «simpatizzante», come dice Stubenberg, del monismo neutrale: afferma infatti che «i rapporti tra mentale e fisico finiranno per essere espressi da una teoria i cui termini fondamentali non potranno essere situati nettamente in nessuna delle due categorie»[31]. Come Spinoza rintraccia la realtà ultima in Dio o nella Natura, Strawson nelle persone,[32] Chalmers nell'informazione e così via, così Nagel rintraccia l'entità neutrale di fondo nel corpo o nel cervello.[33]

---

[28] Cfr. H. Putnam, *Filosofia e vita mentale*, in H. Putnam (1987), pp. 329-330. Afferma qui che «di qualsiasi cosa sia fatta la "stoffa" dell'universo – materiale o spirituale che sia – non coinvolge la libertà perché il mentale è autonomo» (Ivi, p. 330).

[29] Questo fa di Aristotele un monista neutrale? Sia Aristotele che Putnam evitano questa classificazione in quanto il primo non ci permette, a causa dell'oscurità di certi passi, un chiarimento in merito alla sua posizione sulla sostanza ultima (se non come residuo dualistico), mentre il secondo parla di un «mistero» senza però affermare l'inintelligibilità della sostanza ultima. Putnam afferma che «potrebbe anche darsi che, per così dire, la fisica ultima dell'anima sia molto diversa dalla fisica ultima del cervello, ma che al livello che ci interessa, quello dell'organizzazione funzionale, la stessa descrizione vada bene per entrambi» (H. Putnam 1987, p. 329). Possiamo dunque affermare che, almeno nel testo del 1973, sia presente la *tendenza* in Putnam a vedere la sostanza ultima come unica ma descrivibile in termini ambivalenti, secondo un'ottica monistico-neutrale.

[30] Nannini si riferisce allo scritto *Thinking causes* (in *Mental causation* a cura di G. Heil e A. Mele, Clarendon Press, Oxford, pp. 3-17), ritenendo Davidson che «gli eventi in sé né fisici né mentali divengono gli uni o gli altri a seconda del modo nel quale sono descritti» (S. Nannini 2002, p. 126).

[31] T. Nagel, *Che cosa si prova ad essere un pipistrello?*, in D. R. Hofstadter-D. C. Dennett (2001), p. 391.

[32] Peter F. Strawson, in *Individui* del 1959, dimostra «la primitività del concetto di persona (da non risolversi per lui in un corpo da un lato e una coscienza dall'altro)» (AA. VV. 1993, alla voce «Strawson»).

[33] Cfr. L. Stubenberg (2005), § 9.3.

Due bioingegneri italiani dell'Università di Genova, Riccardo Manzotti e Vincenzo Tagliasco, affermando che «la cultura scientifica odierna è dichiaratamente antimetafisica; ma, come diceva Aristotele, chi rifiuta la metafisica è costretto a farla»[34], hanno elaborato una teoria della coscienza che presenta molti punti di contatto con la posizione che stiamo trattando. Essi affermano infatti che la coscienza non è riducibile che a se stessa,[35] che il mondo materiale non è autonomo ma necessita di un osservatore cosciente affinché prenda forma,[36] e dunque che è necessario tenere sempre unite l'epistemologia e l'ontologia perché non si può distinguere ciò che si conosce da ciò che è.[37] L'«esistenza» stessa, come la «rappresentazione» e l'«essere in relazione-con» sono manifestazioni di un'unica realtà fondamentale: l'*onfene*. Formata dalle parole «*on*tologia», «*fen*omenologia» ed «*e*pistemologia», l'onfene è detta anche *relazione intenzionale*, «sintesi di rappresentazione, relazione ed esistenza»[38], e questa è la realtà ultima: «a un livello elementare della realtà è impossibile che esista qualcosa che possieda una di queste tre proprietà senza possedere le altre»[39].
Seguono difatti il monismo neutrale di Russell, affermando che «anche per Russell gli eventi, e non le sostanze permanenti, sono il materiale di partenza di cui è costituito il mondo»[40], divergendo dal filosofo su alcuni punti.[41]

Certamente il monismo neutrale affascina per la sua abilità nel districarsi a un tempo dal fisicalismo e dal mentalismo, aiutato dall'astrattezza della fisica moderna che rende difficile cogliere in modo unitario l'entità naturale comune della realtà. Per questi motivi si sono accostati ad esso pensatori come Mach e Russell, anche se le obiezioni mossegli sono altrettanto forti. La prima obiezione critica le entità neutrali ritenendole mentali (Lenin, Popper e

[34] R. Manzotti-V. Tagliasco (2001), p. 12.

[35] Ivi, p. 65.

[36] Ivi, p. 79.

[37] Ivi, p. 188.

[38] Ivi, p. 205.

[39] Ibidem. Aggiungono gli autori: «il mondo è un insieme di relazioni intenzionali legate in un flusso perenne di intenzionalità» (Ivi, p. 256). Il rapporto mente-corpo è così spiegato, per cui «la continuità della nostra coscienza è una continuità di matrice intenzionale e non fisica o chimica» (Ivi, p. 475).

[40] Ivi, p. 517.

[41] Per esempio: «i) i suoi percetti non hanno un particolare criterio aggregante che permetta di decidere quando un percetto faccia parte di un soggetto, ii) i suoi percetti sono entità che si muovono nello spazio tempo (che a essi preesiste) e devono coabitare la stessa dimensione delle particelle atomiche, iii) i percetti di Russell restano riduzionisticamente divisi, sono autosufficienti» (Ivi, p. 518).

Eccles, Nagel, Chalmers), la seconda si fonda sull'agnosticismo di Nagel e Lockwood ritenendo «intrattabili» certe affermazioni di Russell, la terza evidenzia l'incapacità di distinguere la realtà dalle «illusioni dei sensi»[42] (distinzione che in Russell non può esistere).[43] Il monismo neutrale rappresenta comunque, nell'attuale contesto filosofico-scientifico, la più percorribile «terza via».

Come visto a proposito del monismo neutrale, le vie che sembrano presentarsi sono due: *a*, ridurre la mente e la coscienza all'interno dell'attuale concezione scientifica della natura (naturalizzare la mente); *b*, ampliare il concetto di natura fino ad includervi la coscienza e la fenomenicità dell'esperienza, anche se non attualmente indagabili (antiriduzionismo). La prima via induce a concepire la coscienza come prodotto dell'evoluzione biologica, mentre la seconda ammette l'inspiegabilità del fenomeno coscienziale con l'intento di rivedere l'intero concetto di natura rifondando le scienze naturali. Probabilmente la scienza è «incompleta» come Nagel e Jackson ritengono ma ciò non esclude che la mente sia un prodotto dell'evoluzione. Condividendo la tesi di W. Sellars, per cui il pensiero è un prodotto del linguaggio a sua volta prodotto dell'interazione sociale,[44] sembra plausibile intraprendere la seguente via: nell'ottica dell'evoluzione della specie *Homo*, l'adattamento ambientale e la selezione genica hanno condotto la nostra specie alla formazione *naturale* di, rispettivamente al tempo, *comunicazione pre-linguistica* (non verbale, comune ad altri animali), *coscienza* (intesa come la consapevolezza di esistere e percepire, comune probabilmente a poche altre specie animali) e *linguaggio* (verbale, tipico della specie *Homo Sapiens*). Dunque la mente (il pensiero cosciente e inconscio) è sì un prodotto della selezione naturale ma ciò non significa che si può parlare di «identità mente-cervello». Questa identità confinerebbe la mente alla struttura biologica cerebrale e tuttavia il cervello è necessario ma non suffi-

[42] L. Stubenberg (2005), § 7.5. Nel testo: «illusions of sense».

[43] Cfr. Ivi, cap. 7. F. H. Bradley segnala la conflittualità del monismo neutrale con l'idealismo da lui proposto: «se anima e corpo debbono, in ultima analisi, essere una cosa sola, allora si finisce con l'ammettere una pluralità di reali finiti entro l'Assoluto […] non riesco assolutamente a scorgere per quale motivo la specifica natura di ognuna di queste due forme dovrebbe conservarsi come tale» (F. H. Bradley 1984, p. 478).

[44] Cfr. W. Sellars (2004), §§ 56-59. Espone tesi similare anche Heidegger in *In cammino verso il linguaggio* del 1960. Contrario è invece Edelman (cfr. G. M. Edelman 1993b, p. 195), il quale concepisce il linguaggio come ponte tra coscienza primaria e coscienza d'ordine superiore (cfr. Ivi, p. 207); cfr. G. M. Edelman-G. Tononi (2000), pp. 233-241. Contrari sono anche Manzotti e Tagliasco (cfr. R. Manzotti-V. Tagliasco 2001, pp. 54-57, pp. 312-317 e in particolare p. 364).

ciente per una totale comprensione della mente: il linguaggio ha fatto sì che la relazione con l'ambiente costituisse i processi mentali, ma a loro volta questi ricostituiscono (come per feedback) il collegamento con l'ambiente, venendosi così a creare un circolo (ambiente – cervello/corpo – ambiente) che costituisce la *mente*. L'idea che parte della mente sia costituita dall'ambiente – inteso anche come complesso di cultura e informazione, sempre in evoluzione - appartiene a Gregory Bateson e può ricondursi ad Heidegger.[45] Bateson parla di «circolarità cibernetiche dell'io e del mondo esterno»[46], dove con «cibernetica» si intende quell'ambiente naturale e artificiale in cui domina il principio dell'auto-correzione e della retroazione, per cui «la mente è immanente in quei circuiti cerebrali che sono interamente contenuti nel cervello [oppure] nei circuiti che sono interamente contenuti nel sistema: cervello *più* corpo; [oppure] nel più vasto sistema: uomo *più* ambiente»[47]. Questa circolarità presuppone che «il mondo mentale […] non è delimitato dall'epidermide»[48] e soprattutto

[45] I punti di contatto tra i due pensatori è ben messo in luce da Eugenio Mazzarella in *Un'appendice ontica: ecologia e cibernetica e circolo linguistico. Bateson e Heidegger. La Frömmigkeit come compito ineludibile del pensiero*, in E. Mazzarella (2002), pp. 337-357. In termini differenti, Bateson e Heidegger affermano che «la finalità cosciente ha ora il potere di turbare gli equilibri del corpo, della società e del mondo biologico intorno a noi» (G. Bateson, *Finalità cosciente e natura*, in G. Bateson 1976, p. 474). Entrambi parlano di «saggezza»: «di una coscienza» per Bateson, quella «che si sa radicata nel circolo ermeneutico del disvelamento» (E. Mazzarella 2002, p. 351) per Heidegger. Mazzarella afferma, sulla scorta dei testi batesoniani e heideggeriani, che «si tratta come si vede di conseguire una ridefinizione dell'atteggiamento dell'uomo nell'interazione con il suo ambiente all'altezza della tecnica di cui *già* dispone» (Ivi, p. 347). Ammonisce Bateson infatti: «la carenza di saggezza sistemica è sempre punita» (G. Bateson 1976, p. 474).

[46] G. Bateson, *Effetti della finalità cosciente sull'adattamento umano*, in Ivi, p. 485.

[47] G. Bateson, *La cibernetica dell''io': una teoria dell'alcolismo*, in Ivi, p. 366. Egli afferma che «l'unità di sopravvivenza è il complesso flessibile organismo-nel-suo-ambiente» (Ivi, p. 491) e che «non ha più senso ipotizzare un dualismo tra mente e materia, […] la rete pensante non è confinata al corpo, ma comprende tutti i canali di informazione, in entrata e in uscita» (M. Groppo-A. Antonietti 1992, p. 273). Concetto simile si ritrova in Chalmers: la «mente estesa» è costituita solo in parte dalla mente che funge da «barriera di contatto» intermedio «tra l'ambiente fisico e l'universo della conoscenza» (S. Tagliagambe 2004, p. 191). Si può accostare al pensiero batesoniano anche l'approccio di Francisco Varela (recentemente scomparso) che, insieme a H. Maturana, concepisce «i processi mentali come dei fenomeni estremamente complessi che si situano in una sorta di ambito di confine tra ciò che è cognitivo e ciò che è biologico» (M. Groppo-A. Antonietti 1992, p. 277); essi propongono infatti la teoria dell'«autopoiesi» che caratterizzerebbe ogni sistema vivente come «un'unità [omeostatica] composita di interazioni, collocata in un ambiente con cui interagisce» (Ivi, p. 276).

[48] G. Bateson, *Forma, sostanza e differenza*, in G. Bateson (1976), p. 495. Cfr. inoltre Ivi, p. 499e p. 502.

un'ottica in cui la «mente individuale» è un «sottosistema» della «Mente» definita come «unità immanente nel grande sistema biologico, l'ecosistema»[49].

Il fatto che la mente sia un prodotto naturale dunque non significa che una intelligenza artificiale sia possibile.[50] Condividendo il punto di vista heideggeriano secondo cui «il pensiero calcolante non è il pensiero meditante, il pensiero che contempla il significato che regna in tutto ciò che è»[51], per quanto concerne l'I.A. si può affermare che: *a*, ogni teoria computazionale (funzionalismo e cognitivismo) *confonde il cervello per una calcolatrice*; ovvero il funzionamento «logico» di un cervello non è algoritmico – in nessun senso, cioè in qualsiasi modo l'algoritmo venga concepito – mentre ogni teoria computazionale prevede sempre un algoritmo, frutto di una «intuizione» tipicamente umana, che non ha alcunché di logico; *b*, il progetto di *creare* l'intelligenza artificialmente si può considerare solamente un *vezzo della scienza*, poiché senza alcuna utilità né teorica né pratica.

Riguardo il punto *a*, è probabile che l'Intelligenza Artificiale (come progetto) non possa raggiungere l'intelligenza umana ma solo quella dell'«animale» heideggeriano: per il filosofo tedesco questo può avere comportamento «intelligente» ma non «un rapporto com-

---

[49] Ivi, p. 501. Nella pagina successiva Bateson parla di un'*ecologia delle idee* tratta dagli studi di G. Vickers, paragonabile alla recente teoria di R. Dawkins dei «memi», unità mentali e culturali che agiscono come i geni: per riproduzione. A proposito della coscienza in senso ampio riferita alle macchine, Bateson afferma efficacemente che «ogni ulteriore passo verso un aumento di coscienza porterà il sistema più lontano dalla coscienza totale. Aggiungere la descrizione degli eventi in una certa parte della macchina farà in realtà *diminuire* la percentuale di tutti gli eventi descritti» (G. Bateson, *Finalità cosciente e natura*, in Ivi, p. 472), perché la descrizione di ogni parte del processo complessivo richiede ulteriori circuiti, con un *regressus* all'infinito. Si può considerare un limite strutturale della presa di coscienza di una *parte* nei confronti della coscienza del *tutto*. Dello stesso parere sembrano essere anche Manzotti e Tagliasco quando affermano: «il soggetto non è chiuso nel proprio cervello ma è parte del mondo, ogni trasferimento da un luogo all'altro, modifica la sua mente» (R. Manzotti-V. Tagliasco 2001, p. 273). Dello stesso avviso risulta la parafrasi di Hofstadter del detto cartesiano: «penso; quindi non ho accesso al livello nel quale sommo» (D. R. Hofstadter 1984, p. 733), chiarendo il rapporto, messo in evidenza anche da Bateson, tra conscio e inconscio.

[50] Si intende, a partire da qui, avallare determinate posizioni in materia tenendo a mente il monito di Putnam, secondo cui, «allo stato attuale delle nostre conoscenze, mal converrebbe a chiunque di avere certezze su qualunque posizione relativa ad un mistero così fondamentale come la relazione tra corpo e anima» (H. Putnam 1987, p. 18).

[51] C. Mitcham, *Philosophy of information technology*, in L. Floridi (2004), p. 334. Nel testo: «calculative thinking is not meditative thinking, not thinking which contemplates the meaning which reigns in everything that is».

prendente verso gli oggetti»[52] perché né il suo ambiente né esso stesso «sono manifesti in quanto enti»[53]. All'animale, come ad ogni intelligenza artificiale, manca il *mondo*. In questo senso è condivisibile la posizione di Searle, in particolare in merito alla biologicità del fenomeno mentale e al tempo stesso alla sua irriducibilità ai processi neurobiologici (nel senso di cui parlano Dennett e Hofstadter, per cui al livello neurale non si ha alcuna coscienza mentre questa «appare» soltanto «salendo di livello», cioè al livello dell'individuo che si sente non una «colonia» ma un'unità)[54] oltre che per le critiche al funzionalismo.[55] Proprio a seguito di quest'ultime, Searle è stato criticato da Dennett e Hofstadter con due fondamentali obiezioni, del tutto condivisibili (le uniche peraltro): 1. il «potere causale» di cui Searle parla, che unicamente nel cervello darebbe vita alla mente, è un concetto troppo vacuo; 2. con esplicito riferimento all'esperimento mentale della «stanza cinese», "tradurre" una lingua integrandosi con le regole di traduzione – cioè "conoscendole", in base all'«obiezione dei sistemi» – corrisponde a "capirla" (l'uomo «sa» le regole e il calcolatore no, mentre se le «sapesse» capirebbe come l'uomo). Una probabilmente valida risposta a tali obiezioni sarebbe la seguente. Una parola viene «capita» perché non è solo una configurazione di scariche neurotiche. Se così fosse allora saremmo macchine non pensanti ma calcolanti. Il calcolatore (ed è qui l'errore dell'obiezione) possiede solo il livello corrispondente alle scariche elettriche, cui *noi* assegniamo i valori di «0» e «1», e livelli che corrispondono solo a costrutti logici che *noi* interpretiamo come tali: la macchina registra solo scariche elettriche (la cui eventuale assenza siamo *noi* ad interpretare). Solo l'intelligenza umana è in grado di «salire di livello» in quanto l'evoluzione ha permesso una strutturazione del

[52] C. Di Martino (2002), p. 98.

[53] Ivi, p. 99, dove riprende lo stesso Heidegger. Heidegger ritiene anche biologicamente diversi uomo (unità di organismo, anima, spirito e «estensivo») e animale ma questo è difficilmente condivisibile, al contrario del ritenere quanto sia necessario «*abitare* consapevolmente la nostra differenza […] della parola e della morte […] in cui ci troviamo per così dire gettati, nel riconoscimento che è precisamente nello spazio di tale differenza che si gioca (anche) il destino di ogni vivente» (Ivi, p. 122).

[54] Cfr. D. R. Hofstadter (1984), p. 378.

[55] Potrebbe non risultare condivisibile l'identificazione, compiuta da Searle, di mente e coscienza. Inoltre risulta difficile concepire «come» l'I.A. possa riuscire, anche concettualmente, a superare gli ostacoli messi in luce dallo stesso Searle – il più «forte» dei quali è la nozione di «ruolo causale intenzionale» che solo il cervello può possedere – se il suddetto «ruolo causale» non può far parte del mondo artificiale. Lo stesso tipo di contraddizione si può rintracciare in Edelman, che elabora teorie anti-funzionaliste per proporne una connessionista (come se si aggirassero ostacoli per poi cadere in altri simili).

cervello tale che un impulso esterno (per esempio – attraverso il canale uditivo – un suono), al momento di giungere al cervello viene collegato ad altri impulsi che precedentemente erano stati "registrati" come configurazioni di scariche neuronali (individuali e sempre cangianti). Dunque, tenendo sicuramente in largo conto le scoperte – e le ipotesi – (in particolare quelle compiute da Edelman) delle neuroscienze, è possibile ipotizzare che questo suono giunto al cervello divenga "parola" comprensibile e compresa, *colta*, in virtù del fatto che la complessità del cervello ha prodotto livelli *reali* distinti di *aggregazione di unità*: il suono può essere così "interpretato" dal livello non "sinaptico" ma da uno "superiore" che può essere coscienziale (cosciente) oppure subcoscienziale (inconscio) ma in ogni caso capace di collegare il suono – adesso "parola" – ad altri "ricordi" senza tenere conto della sola configurazione neuronale corrispondente ad esso. Le macchine hanno un solo livello reale e confondere i diversi livelli informatici[56] per livelli reali e con proprietà effettivamente distinte è un errore.[57] Questi livelli sono "virtuali" in quanto presenti "nell'occhio di chi guarda" cioè del programmatore. Il cervello, grazie alla complessità strutturale frutto di milioni di anni di evoluzione, presenta più livelli "reali" e nessuno "virtuale". Il confronto di livelli del tipo neuroni-circuiti, scariche sinaptiche-connessioni elettriche su piastra, cervello-processore di dati e mente-software o struttura logica, è fuorviante. Sembra più reale il seguente tipo di confronto per livelli: *I.* le scariche elettriche della macchina corrispondono alle scariche elettriche del cervello (livello reale per entrambi, unico non fittizio per la macchina); ulteriori

[56] Seguiamo Manzotti e Tagliasco: «al livello più basso i valori assunti dalla tensione elettrica, poi il significato logico (0, 1), poi il livello del codice macchina (Assembler), poi il livello dei linguaggi di programmazione (C, C++, Pascal), infine il livello con il sistema operativo» (R. Manzotti-V. Tagliasco 2001, p. 43). Hofstadter divide analogamente in: linguaggio macchina, assemblatore, compilatore e interprete (D. R. Hofstadter 1984, p. 312).

[57] Così affermano anche Manzotti e Tagliasco: ogni livello, a parte le tensioni elettriche e la mente dell'utente, esiste in rapporto al soggetto cosciente, ma non in sé; è «il frutto della giustapposizione, a un fenomeno fisico, di un determinato schema di significati» (R. Manzotti-V. Tagliasco 2001, p. 43). Cade nell'errore, seconda la prospettiva che qui si avalla, Hofstadter: «il passo cruciale che bisogna compiere è quello che porta da una descrizione a basso livello (cioè neurone per neurone) dello stato di un cervello alla descrizione di alto livello (cioè modulo per modulo) dello stesso stato dello stesso cervello» (D. R. Hofstadter 1984, p. 378); «salendo di livello», una macchina potrebbe così «avere coscienza». Addirittura identifica i seguenti livelli: il linguaggio macchina e le proteine, il programma di alto livello (inteso come lunga sequenza di dati, poi tradotta in linguaggio macchina) e il DNA, il processore centrale (CPU) e la cellula, infine i processori (e interpreti) e ribosomi (e molecole di tRNA). Cfr. Ivi, p. 591.

livelli reali della sola struttura cerebrale: *II*. configurazioni neuronali, *III*. mappe rientranti edelmaniane, *IV*. «potere causale» di Searle (*capacità semantica intenzionale*). La differenza tra macchina e cervello spiega come ogni riferimento comportamentista di Dennett e ogni posizione panpsichista[58] siano completamente fuor di luogo nell'ambito del mentale, della coscienza e della *semanticità* della macchina uomo: non esiste macchina non umana che possa pensare, essere cosciente e *comprendere*.[59]

Riguardo invece il punto *b*, è probabile che l'Intelligenza Artificiale debba occuparsi di capire le capacità cognitive non di un elaboratore o dell'uomo, bensì di esseri viventi, seppur con capacità cognitive ancora avvolte nel mistero, la cui struttura interna è per lo meno molto più semplice, in senso evolutivo, rispetto a quella umana, al fine di render cosciente (come un uomo) un elaboratore di dati. Insomma, prima di tentare di "trovare" la mente, la coscienza e l'intenzionalità nel cervello umano, bisogna tentare di capire se, ed eventualmente come, queste proprietà naturali compaiano in forme di vita più elementari.[60] Così si otterrebbero maggiori risultati e l'impostazione del progetto di ricerca assumerebbe una forma più scientifica e metodica: *studio comparato delle forme intelligenti*. E l'intelligenza non dovrebbe essere più un parametro misurabile con i test di tipo comportamentista come quello di Turing ma dovrebbe essere considerata alla stregua della «comprensione», proprio mettendo in luce una probabile etimologia del termine: «intus-legere»[61] o «unire dentro», «collegare», «connettere», ovvero capire, «cogliere» il *senso* della realtà. Importante è però capire quanto la ricerca di una intelligenza artificiale non sia da considerare utile alla scien-

---

[58] Il riferimento è diretto a Dennett e Hofstadter che attribuiscono una mente a praticamente ogni struttura che presenti una certa complessità (probabilmente tutto), per seguire il quesito comportamentista «come faccio a sapere che sei cosciente?» per cui il robot è cosciente nella misura in cui la coscienza viene «convenzionalmente» attribuita a chi mi circonda. Ma una calcolatrice ha «un particolare e "soggettivo" punto di vista»? Per questo motivo sembra fuor di luogo ogni discussione di tipo panpsichista. Cfr. D. R. Hofstadter-D. C. Dennett (2001), p. 398.

[59] L'importanza del significato è ribadita da Manzotti e Tagliasco, per i quali il concetto basilare per la ricerca dell'I.A., ovvero l'informazione, non essendo autonomo rispetto all'esistenza di osservatori coscienti, dipende da un'attribuzione di *significato* (cfr. R. Manzotti-V. Tagliasco 2001, p. 73 e ss.). Cfr. pure F. Cimatti (2004).

[60] Brooks tenta questo approccio con discreti risultati, vicini alla proposta teorica di Manzotti e Tagliasco.

[61] In G. Devoto (2004), alla voce «intelligente», l'etimologia suggerita è invece «inter-legere» o «saper scegliere tra». L'etimologia qui proposta è intesa a far transitare la terminologia da un campo semantico della *computazione* (quale la scelta tra più soluzioni) a uno dell'*interrelazione di significati* (quale il collegamento interdipendente tra la mente e l'esterno in un processo di interiorizzazione relazionale).

za e dunque all'uomo. La ricerca scientifica dovrebbe essere sempre diretta a concepire nuovi metodi di vita per un migliore adattamento all'ambiente in cui non solo *abitiamo* (anche in senso heideggeriano) ma che, per i motivi visti, in un certo senso *siamo*. Secondo questo orientamento metodologico, il progetto dell'I.A. «forte» appare del tutto inappropriato: esso si dimostra interessato alla «creazione» di una intelligenza al di fuori di quella umana, senza tener conto del beneficio che questa eventuale (quanto, forse, inutile) creazione possa portare all'uomo. Lo sviluppo di una macchina pensante quale benefici può portare all'umanità? Nessuna.[62] Ma va considerato l'altro aspetto della questione.[63] Le macchine hanno infatti «aperto i nostri occhi» alla conoscenza di noi stessi, in un modo nuovo nella storia umana, in quel processo di autoanalisi che l'uomo ha sempre operato. Nel confronto con la macchina, l'uomo ha potuto conoscere meglio la mente e sicuramente in tal senso – e forse attualmente solo in questo – l'I.A. è degna di un ruolo scientifico: il modo di «funzionare» di una macchina rivela quanto gli umani «non funzionino» ma «pensino», non «sussistano» ma «vivano», non «calcolino» ma «riflettano», non «stampino» ma «comunichino», non «eseguano» ma «scelgano liberamente».

Il progetto dell'Intelligenza Artificiale ha il pregio di *connotare* la specie umana, nei suoi difetti, che *deve estirpare*, e nelle sue peculiari doti, che *deve mettere a frutto*.

[62] Le ricerche in campo neuroscientifico sul cervello (aspetto essenziale della ricerca nel campo dell'I.A.) sono utilissime alla comprensione dell'universo e del suo funzionamento, necessarie nella concezione di «utilità scientifica» qui discussa.

[63] Putnam rintraccia una nota positiva nell'I.A., la nozione di organizzazione funzionale, e una negativa, la sua eccessiva semplificazione (cfr. H. Putnam, *Filosofia e vita mentale*, in H. Putnam 1987, p. 327).

Riferimenti bibliografici[1]

AA. VV.

[1993] *Enciclopedia Garzanti di Filosofia*, Edizione Mondolibri su licenza Garzanti, Milano

Andolfo, M. (a cura di)

[1999] *Atomisti antichi. Testimonianze e frammenti*, Rusconi, Milano

Arendt, H.

[1987] *La vita della mente*, trad. di G. Zanetti e introduzione di A. Dal Lago, Il Mulino, Bologna

Bailey, J.

[1996] *Il postpensiero. La sfida dei computer all'intelligenza umana*, trad. di L. Sosio, Garzanti, Milano

Bateson, G.

[1976] *Verso un'ecologia della mente*, trad. di G. Longo e G. Trautteur, Edizione Mondolibri su licenza Adelphi, Milano

Bechtel, W.

[1992] *Filosofia della mente*, trad. di M. Salucci, Il Mulino, Bologna

Bettelli, O.

[2004] *La mente indaga la mente*, Starrylink, Brescia

Biuso, A. G.

[2004] *Cyborgsofia. Introduzione alla filosofia del computer*, Il pozzo di Giacobbe, Trapani

Boniolo, G., **Vidali, P.**

[2003] *Introduzione alla filosofia della scienza*, Bruno Mondadori, Milano

Bradley, F. H.

[1984] *Apparenza e realtà. Saggio di metafisica*, introduzione e trad. di D. Sacchi, Rusconi, Milano

Brockman, J.

[1999] *La terza cultura. Oltre la rivoluzione scientifica*, trad. di L. Carra, Garzanti, Milano

[1] Per l'anno di edizione si è preferito rimandare all'anno dell'edizione italiana, non di quella originale (per i testi di origine straniera pubblicati in Italia), e comunque sempre all'ultima edizione.

Burattini E., **Cordeschi R.** (a cura di)

[2001] *Intelligenza artificiale. Manuale per le discipline della comunicazione*, Carocci, Roma

Capucci, P. L. (a cura di)

[1994] *Il corpo tecnologico. L'influenza delle tecnologie sul corpo e sulle sue facoltà*, Baskerville, Bologna

Carli, E. (a cura di)

[2003] *Cervelli che parlano. Il dibattito su mente, coscienza e intelligenza artificiale*, Bruno Mondadori, Milano

Ceruti, M., **Preta, L.** (a cura di)

[1990] *Che cos'è la conoscenza*, Laterza, Roma-Bari

Changeux, J. P.

[1983] *L'uomo neuronale*, trad. di C. Sughi, Feltrinelli, Milano

Cimatti, F.

[2004] *Il senso della mente. Per una critica del cognitivismo*, Bollati Boringhieri, Torino

Civita, A.

[2003] *Filosofia della psicologia*, in Vassallo (2003b), pp. 281-318

Crick, F. H. C.

[1979] *Riflessioni sul cervello*, articolo in *Le Scienze* n. 135, Novembre 1979, pp. 160-168

D'Agostini, F., **Vassallo, N.** (a cura di)

[2002] *Storia della filosofia analitica*, Einaudi, Torino

Damasio, A. R.

[1995] *L'errore di Cartesio. Emozione, ragione e cervello umano*, trad. di F. Macaluso, Adelphi, Milano

Dennett, D. C.

[1997] *La mente e le menti. Verso una comprensione della coscienza*, trad. di I. Blum, Rizzoli, Milano

Devoto, G.

[2004] *Dizionario etimologico. Avviamento alla etimologia italiana*, Edizione Mondolibri su licenza Le Monnier, Milano

Di Francesco, M.

[1996] *Introduzione alla filosofia della mente*, Carocci, Roma

[2000] *La coscienza*, Laterza, Roma-Bari

Di Martino, C.

[2002] *L'uomo e l'animale, la morte e la parola*, in Leoni, F., Maldonato, M. (a cura di), *Al limite del mondo. Filosofia, estetica, psicopatologia*, Dedalo, Bari, pp. 89-123

Di Spazio, V.

[2004] *Cronobiocibernetica o del tempo somatizzato*, articolo presente sul sito internet «www.biuso.it/testi.htm»

Dorato, M.

[2003] *Filosofia della fisica*, in Vassallo (2003b), pp. 99-140

Du Sautoy, M.

[2004] *L'enigma dei numeri primi. L'ipotesi di Riemann, l'ultimo grande mistero della matematica*, trad. di C. Capararo, Edizione Mondolibri su licenza RCS Libri, Milano

Eco, U.

[2002] *Kant e l'ornitorinco*, Bompiani, Milano

Edelman, G. M.

[1993a] *Topobiologia. Introduzione all'embriologia molecolare*, trad. di G. Corbellini e D. Ribatti, Bollati Boringhieri, Torino

[1993b] *Sulla materia della mente*, trad. di S. Frediani, Adelphi, Milano

[1995] *Darwinismo neurale. La teoria della selezione dei gruppi neuronali*, trad. di S. Ferraresi e prefazione di G. Tononi, Einaudi, Torino

[2004] *Più grande del cielo. Lo straordinario dono fenomenico della coscienza*, trad. di S. Frediani, Einaudi, Torino

Edelman, G. M., **Tononi, G.**

[2000] *Un universo di coscienza. Come la materia diventa immaginazione*, trad. di S. Ferraresi, Einaudi, Torino

Eraclito

[2005] *Dell'Origine*, trad. e cura di A. Tonelli, Feltrinelli, Milano

Farina, B.

[1999] *Dal* Ricordo del Presente *di Bergson al* Presente Ricordato *di Edelman*, Franco Angeli, Milano

Fimiani, M., **Gessa Kurotschka V.**, **Pulcini E.** (a cura di)

[2004] *Umano post-umano. Potere, sapere, etica nell'età globale*, Editori Riuniti, Roma

Floridi, L. (edited by)

[2004] *The Blackwell guide to the philosophy of computing and information*, Blackwell, Oxford

Fodor, J. A.

[2001] La *mente non funziona così. La portata e i limiti della psicologia computazionale*, trad. di M. Marraffa, Laterza, Roma-Bari

Frixione, M.

[1994] *Logica, significato e intelligenza artificiale*, Franco Angeli, Milano

[2003] *Filosofia delle scienze cognitive*, in Vassallo (2003b), pp. 319-350

Gardner, M.

[1979] *«*Gödel, Escher, Bach*» di Douglas R. Hofstadter*, recensione in *Le Scienze* n. 135, Novembre 1979, pp. 170-174

Gazzaniga, M. S.

[1990] *Stati della mente stati del cervello. Come interagiscono mente e cervello per creare la nostra vita cosciente*, trad. di G. Noferi, Giunti, Firenze

Gava, G.

[1983] *Il problema mente cervello. Genesi e sviluppo della teoria dell'identità*, prefazione di V. Somenzi, Cortina, Milano

Gessa Kurotschka, V.

[2004] *I saperi del vivente e la filosofia*, in Fimiani-Gessa Kurotschka-Pulcini (2004), pp. 117-136

Giaretta, P.

[2003] *Filosofia della logica*, in Vassallo (2003b), pp. 141-178

Goleman, D.

[1999] *Intelligenza emotiva*, trad. di I. Blum e B. Lotti, Rizzoli, Milano

Groppo, M., **Antonietti, A.** (a cura di)

[1992] *Nuove teorie della mente. Concezioni recenti su mente, pensiero, intelligenza*, Vita e Pensiero, Milano

Hendry, R. F.

[2003] *Filosofia della chimica*, in Vassallo (2003b), pp. 39-70

Hofstadter, D. R., **Dennett, D. C.**

[2001] *L'io della mente. Fantasie e riflessioni sul sé e sull'anima*, trad. di G. Longo, Adelphi, Milano

Hofstadter, D. R.

[1984] *Gödel, Escher, Bach: un'Eterna Ghirlanda Brillante. Una fuga metaforica su menti e macchine nello spirito di Lewis Carroll*, trad. di B. Veit, G. Longo, G. Trautteur, S. Termini e B. Garofalo, Adelphi, Milano

Horgan, J.

[1998] *La fine della scienza*, trad. di T. Cannillo, Adelphi, Milano

Johnson-Laird, P. N.

[1988] *La mente e il computer. Introduzione alla scienza cognitiva*, trad. di P. Tabossi, Il Mulino, Bologna

Kant, I.

[1995] *Prolegomeni ad ogni metafisica futura che vorrà presentarsi come scienza*, trad. di P. Martinetti, Rusconi, Milano

[2002] *Critica della ragion pura*, trad. di G. Gentile e G. Lombardo-Radice, Edizione Mondolibri su licenza Laterza, Milano

Lavazza, A.

[2001] *La mente? Si è allargata*, recensione in *Avvenire*, 5 Ottobre 2001

[2004] *La mente allargata*, articolo in *Mente & Cervello* n. 11, Settembre-Ottobre 2004, pp. 60-63

Legrenzi, P.

[2002] *La mente*, Il Mulino, Bologna

Lenzen, M.

[2004] *Una macchina con l'anima*, articolo in *Mente & Cervello* n. 12, Novembre-Dicembre 2004, pp. 88-91

Leoni, F.

[2002] *L'inappropriabile*, in Leoni, F., Maldonato, M. (a cura di), *Al limite del mondo. Filosofia, estetica, psicopatologia*, Dedalo, Bari, pp. 67-87

Lis, A., **Stella, S.**, **Zavattini, G. C.**

[1999] *Manuale di psicologia dinamica*, Il Mulino, Bologna

Losee, J.

[2001] *Filosofia della scienza. Un'introduzione*, trad. di P. Budinich, Il Saggiatore, Milano

Lurija, A. R.

[1992] *Un mondo perduto e ritrovato*, trad. di S. Arcella e prefazione di O. Sacks, Editori Riuniti, Roma

Manzotti, R., **Tagliasco, V.**

[2001] *Coscienza e realtà. Una teoria della coscienza per costruttori e studiosi di menti e cervelli*, Il Mulino, Bologna

Marconi, D.

[2001] *Filosofia e scienza cognitiva*, Laterza, Roma-Bari

Marraffa, M.

[2002] *Filosofia della mente*, in D'Agostini-Vassallo, pp. 194-222

Mazzarella, E.

[2002] *Tecnica e metafisica. Saggio su Heidegger*, Guida, Napoli

Minsky, M.

[1989] *La società della mente*, trad. di G. Longo, Adelphi, Milano

Modeo, S.

[2004] *In tre mosse nasce la coscienza*, recensione in *Darwin* n. 4, Novembre-Dicembre 2004, p. 58

Moriconi, E.

[2003] *Filosofia della matematica*, in Vassallo (2003b), pp. 179-212

Nannini, S.

[2002] *L'anima e il corpo. Un'introduzione storica alla filosofia della mente*, Laterza, Roma-Bari

Negrotti, M. (a cura di)

[1990] *Capire l'artificiale. Dall'analogia all'integrazione uomo-macchina*, Bollati Boringhieri, Torino

O'Connor, T., **Robb, D.** (edited by)

[2003] *Philosophy of mind. Contemporary readings*, Routledge, London

Odifreddi, P.

[2003] *Il diavolo in cattedra. La logica da Aristotele a Gödel*, Edizione Mondolibri su licenza Einaudi, Milano

Oliverio, A.

[2004] *Miti e realtà delle neuroscienze*, in Fimiani-Gessa Kurotschka-Pulcini (2004), pp. 137-156

Penrose, R.

[1997] *La mente nuova dell'imperatore. La mente, i computer e le leggi della fisica*, trad. di L. Sosio e prefazione di M. Gardner, Rizzoli, Milano

Popper, K. R.

[1996] *La conoscenza e il problema corpo-mente*, trad. di F. Laudisa, Il Mulino, Bologna

Pratt, V.

[1990] *Macchine pensanti. L'evoluzione dell'intelligenza artificiale*, trad. di M. T. Bolla e M. L. Sapino, Il Mulino, Bologna

Putnam, H.

[1987] *Mente, linguaggio e realtà*, trad. di R. Cordeschi, Adelphi, Milano

[1993a] *Matematica, materia e metodo*, trad. di G. Criscuolo, Adelphi, Milano

[1993b] *Rappresentazione e realtà*, trad. di N. Guicciardini, Garzanti, Milano

Reale, G.

[2002] *Introduzione a Aristotele*, Laterza, Roma-Bari

[2004] *Storia della filosofia greca e romana. Volume 1. Orfismo e presocratici naturalisti*, Bompiani, Milano

Reale, G., **Antiseri, D.**

[1998] *Il pensiero occidentale dalle origini a oggi. Volume 3. Dal Romanticismo ai giorni nostri*, La Scuola, Brescia

Rose, S.

[1994] *La fabbrica della memoria. Dalle molecole alla mente*, trad. L. Sosio, Garzanti, Milano

Russell, B.

[1970] *L'analisi della mente*, trad. di J. Sanders e L. Breccia, Newton&Compton, Roma

Sacks, O.

[2004] *Nel fiume della coscienza*, articolo in *Mente & Cervello* n. 11, Settembre-Ottobre 2004, pp. 46-59

Sanford, A. J.

[1992] La *mente dell'uomo*, trad. di M. Riccucci, Il Mulino, Bologna

Searle, J. R.

[1994] *La riscoperta della mente*, trad. di S. Ravaioli, Bollati Boringhieri, Torino

[1998] *Il mistero della coscienza*, trad. di E. Carli, Raffaello Cortina Editore, Milano

Sellars, W.

[2004] *Empirismo e filosofia della mente*, trad. di E. Sacchi, introduzione di R. Rorty e «guida al testo» di R. Brandom, Einaudi, Torino

Shreeve, J.

[2005] *I segreti della mente*, articolo in *National Geographic* n. 3, Marzo 2005, pp. 32-61.

Stubenberg, L.

[2005] *Neutral monism*, monografia elettronica della Stanford Encyclopedia of Philosophy, presente sul sito internet «www.seop.leeds.ac.uk/entries/neutral-monism»

Tabossi, P.

[1988] *Intelligenza naturale e intelligenza artificiale*, Il Mulino, Bologna

Tagliagambe, S.

[2004] *Mente e cervello: un rapporto problematico*, in Fimiani-Gessa Kurotschka-Pulcini (2004), pp. 157-194

Tononi, G.

[2004] *Matematica e geometria della coscienza*, articolo in *Darwin*, n. 4, Novembre-Dicembre 2004, pp. 36-42

Ubaldo, N.

[2003] *Atlante illustrato di Filosofia*, Demetra (Gruppo Giunti), Firenze

Vallar, G.

[1981] *L'io e il suo cervello*, recensione in *Le Scienze* n. 159, Novembre 1981, pp. 151-152

Vassallo, N.

[2002] *Epistemologia*, in D'Agostini-Vassallo, pp. 250-274

[2003a] *Teoria della conoscenza*, Laterza, Roma-Bari

[2003b] *Filosofie delle scienze*, Einaudi, Torino (a cura di)

Walsh, D. M.

[2003] *Filosofia della biologia*, in Vassallo (2003b), pp. 3-38

Williams, S.

[2003] *Storia dell'intelligenza artificiale. La battaglia per la conquista della scienza del XXI secolo,* trad. di A. Antonini, Garzanti, Milano