

INTRODUZIONE

Il soggetto del presente lavoro è una recente ipotesi della scienza cognitiva post-classica che verte sulla natura delle relazioni che intercorrono tra agente e ambiente.

La Mente Estesa è un approccio di ricerca la cui data di nascita viene fatta risalire alla pubblicazione dell'articolo: “*The Extended Mind*” di Andy Clark e David Chalmers, comparso per la prima volta nel 1998 sulla rivista *Analysis*.

Secondo il punto di vista degli autori che avanzano tale ipotesi, la mente non sarebbe delimitata dai confini di cranio e pelle ma, essendo l'agente integrato in un ambiente fisico e sociale, si estenderebbe al mondo esterno. L'uomo è infatti capace di modificare il mondo secondo i suoi bisogni, con azioni epistemiche, e, al tempo stesso, il mondo determina, influenza o semplicemente aiuta l'agente nei suoi processi cognitivi. Tale rapporto bidirezionale si configura come aperto e del tutto naturale.

La mente “emerge”, dunque, dall'accoppiamento (*coupling*) di cervello, corpo e mondo.

Uno degli esempi più classici forniti dai sostenitori della teoria è quello della moltiplicazione con carta e penna. Pochissime persone al mondo riescono a calcolare “a mente” una moltiplicazione complessa come (4234×4567) ; tutti sono, invece, capaci di farlo con l'ausilio di carta e penna. Questo perché tali “impalcature esterne” ci permettono di scaricare, di volta in volta, parti del processo di elaborazione e dati parziali del calcolo e di alleggerire, quindi, il nostro compito. In questo senso, la mente è estesa a carta e penna.

L'ipotesi della Mente Estesa è da inquadrare, da un lato, nella critica più ampia all'impianto concettuale della cosiddetta scienza cognitiva classica e, dall'altro, nella discussione sulla natura della relazione tra mente e mondo. Il

dibattito di riferimento si concentra, dunque, attorno alla spiegazione dei rapporti che intercorrono tra un agente cognitivo e l'ambiente (fisico, sociale, culturale) nel quale è inserito. La preoccupazione di chi difende l'ipotesi della *Mente Estesa* è che le precedenti teorie cognitive hanno tralasciato un livello di spiegazione più ampio che vede l'uomo e il mondo influenzarsi in un rapporto attivo e dinamico. In particolare, si sostiene che i processi cognitivi e gli stati mentali in generale siano causati, almeno in parte, da elementi esterni al cervello e al corpo dell'agente.

Obiettivo del presente lavoro è quello di dare un quadro di riferimento all'ipotesi della *Mente Estesa*, di descriverne i fondamenti e di discuterne, con l'aiuto delle principali critiche mosse da autorevoli studiosi, la validità.

La struttura della tesi si articola in tre capitoli.

Nel primo si cerca di dare un quadro di riferimento storico generale della scienza cognitiva. Data la vastità del tema di questa sezione (che potrebbe da solo essere oggetto di un'ampia monografia) non si ha pretesa di esaustività ma soltanto di delineare un percorso di riferimento entro il quale collocare l'ipotesi della *Mente Estesa*.

Il secondo capitolo costituisce un'analisi della teoria, la quale viene smembrata in più parti e analizzata nelle sfaccettature individuate. In particolare, dopo aver riportato le caratteristiche generali e i fondamenti teorici, l'indagine segue un percorso che si snoda attorno ai temi di cervello, ambiente, corpo (ovvero i tre principali soggetti del sistema integrato previsto dalla *Mente Estesa*).

Il terzo capitolo riporta, ripercorrendo la stessa struttura del precedente (basi, cervello, ambiente, corpo), le principali critiche e osservazioni mosse alla teoria.

I paragrafi del secondo e del terzo capitolo sono, talvolta, integrati con le considerazioni di chi scrive.

Il titolo di questa tesi ha una sua ragione. *Estesa non mente?* è una domanda la cui interpretazione è lasciata al lettore.

Se da una parte si può abbracciare totalmente l'ipotesi proposta da Clark e Chalmers, ammettendo, quindi, che “la cognizione estesa non mente”, dall'altra parte si può accettare che esiste un rapporto causale bidirezionale tra ambiente e agente pur senza considerarlo parte della mente... in questo caso ciò che è esteso non è mente.

*Il robot non possedeva
sentimenti, solo correnti positroniche che imitavano i sentimenti.
(E forse anche gli esseri umani non avevano sentimenti,
ma solo correnti neuroniche che venivano interpretate come sentimenti...)*
(Isaac Asimov, I robot dell'alba)

*I nostri cervelli rendono il mondo così intelligente
da permetterci di fare i tonti!*
(Andy Clark, Being there)

*Physical concepts are free creations of the human mind,
and are not, however it may seem,
uniquely determined by the external world.*
(Albert Einstein)

CAPITOLO PRIMO

UN'INTRODUZIONE ALLE SCIENZE COGNITIVE

Affinché il nostro studio di caso possa essere meglio compreso, è necessario fare almeno un breve excursus sulle scienze cognitive e sulla loro storia. Questo capitolo delinea una traccia entro la quale collocare la teoria della Mente Estesa e fornisce gli strumenti di confronto all'interno del più vasto contesto dei programmi di ricerca della scienza cognitiva.

1.1 COSA SONO E COME SONO NATE LE SCIENZE COGNITIVE

La scienza cognitiva è un programma di ricerca multidisciplinare che studia i processi cognitivi e le facoltà mentali (percezione, pensiero, ragionamento, pianificazione, problem-solving, decision-making etc), analizzandoli come elaborazioni di un calcolatore finito, naturale o artificiale (Marconi 2001). Vista la multidisciplinarietà del tema, sarebbe forse più appropriato parlare di scienze cognitive al plurale. Di mente e cognizione si occupa, infatti, un rango di discipline che spazia dalla filosofia alla psicologia, dall'intelligenza artificiale alle neuroscienze e dalla linguistica all'antropologia (Marraffa 2003). La data di nascita delle scienze cognitive è individuata attorno al convegno sulla teoria dell'informazione tenutosi al MIT nel dicembre 1956. Si può, comunque, far coincidere tale data di nascita con la creazione dei primi elaboratori che hanno prospettato la possibilità della realizzazione di forme di ragionamento non umano.

La nascita della prima scienza cognitiva vive il passaggio dall'approccio comportamentista a quello cognitivista. Gli assunti base del comportamentismo psicologico si basano sulla negazione, per via della mancanza di oggettività, del metodo introspettivo e affidano la loro analisi

all'osservazione dei comportamenti esterni che vanno individuati in una logica di stimolo-risposta. In questo senso, gli stati mentali non vengono presi in considerazione e la scatola cranica è vista come una *black box* il cui studio appartiene alla sola fisiologia.

Per il comportamentismo filosofico, la mente concepita nei termini della psicologia del senso comune non ha una realtà fisica e gli stati mentali (introdotti come strumento esplicativo per spiegare il comportamento) non sono direttamente identificabili con stati cerebrali.

Tale approccio è dunque ben lontano dal postulare rappresentazioni mentali.

E' proprio ammettendo l'esistenza di strutture e processi interni che, via via, ci si allontana dal comportamentismo per assumere altre visioni e conseguenti metodologie. Gli anni sessanta conoscono lo sviluppo dell'intelligenza artificiale e la commistione degli studi di psicologia e informatica. L'analogia tra mente e macchina era molto forte e si riteneva che i processi cognitivi, paradigmaticamente i processi di ragionamento, potessero essere formalizzati e riprodotti da opportuni programmi per computer.

Il rifiuto dell'impianto concettuale del comportamentismo ha portato alla definizione di un nuovo approccio: il cognitivismo. Esso ammette l'esistenza di strutture interne fondamentali per la spiegazione del comportamento. Tali strutture possono essere studiate dalla psicologia secondo una visione funzionale ovvero in base al ruolo che queste ricoprono nel sistema cognitivo.

La psicologia può quindi "mettere le mani" nella scatola nera ed analizzare quegli stati interni che fino a quel momento erano stati tralasciati. Lo schema stimolo-risposta veniva definitivamente abbandonato. La spiegazione del comportamento richiedeva di postulare rappresentazioni e processi interni che dovevano essere analizzati dallo psicologo cognitivo. Dato che le rappresentazioni non erano considerate mere copie degli stimoli, era

necessario capire che relazione ci fosse tra questi e gli stati interni dell'agente (Paternoster 2002).

Ben presto ci si rese conto che la macchina ha dei grossi limiti e non può competere con la mente in molte prestazioni. Tale consapevolezza è dovuta al presentarsi del “*frame problem*”. La macchina non è in grado di selezionare, di volta in volta, tra le numerose informazioni a cui ha accesso, quelle (più) pertinenti; è limitata a problemi ben definiti; ha bisogno che le conoscenze siano rappresentate esplicitamente e manca di “senso comune” non essendo capace di adattare, in modo flessibile, alle diverse situazioni, i modelli di conoscenza e ragionamento di cui dispone.

La consapevolezza dell'esistenza del *frame problem* è una delle cause della parziale ridefinizione dei metodi e degli obiettivi della scienza cognitiva. Si pone il problema della conoscenza e del ragionamento.

Negli anni ottanta si percepiscono, invece, i limiti dell'intelligenza artificiale. Si affaccia l'ipotesi connessionista, basata sui sistemi neurali, nei quali l'elaborazione viene eseguita in una forma distribuita su più unità che lavorano contemporaneamente. I connessionisti non accettano l'indipendenza della computazione dal supporto materiale (Marconi 2001). Il connessionismo si pone, in tal senso, in opposizione alla cosiddetta scienza cognitiva classica per la quale l'indipendenza della computazione dal supporto materiale è un principio cardine.

In generale possiamo delineare l'avvicinarsi dei contributi della scienza cognitiva di pari passo con lo sviluppo dei sistemi di elaborazione, con l'evoluzione delle teorie psicologiche e filosofiche, con le scoperte neuroscientifiche. Possiamo vedere tali contributi anche secondo l'ottica trasversale del dibattito sulla causazione del mentale che vede la diatriba tra internisti ed esternisti.

In generale, l'internismo (o individualismo) è quella visione secondo la quale esiste una netta distinzione tra gli stati psicologici dell'individuo e l'ambiente fisico e sociale nel quale l'agente è inserito. Carattere distintivo di tale approccio è quello di individuare gli stati psicologici “dentro la testa”. In altri termini, per l'internismo uno stato mentale è completamente specificato da fattori tutti interni alla testa dei soggetti. L'estremizzazione di questa tesi porta al solipsismo (potrebbe esistere un solo individuo e quello che percepisce è creato dalla sua coscienza).

L'esternismo, come avremo ampiamente modo di vedere, contrariamente all'individualismo, riconosce un importante ruolo all'ambiente e sottolinea l'esistenza di uno scambio continuo tra agente e contesto nel quale questi è inserito.

Si può dire che l'esternismo è una teoria più giovane rispetto alla sua opposta. Ad esso fanno, infatti, capo alcune delle più recenti teorie della scienza cognitiva post-classica, tra cui l'ipotesi della Mente Estesa.

1.2. LA SCIENZA COGNITIVA CLASSICA

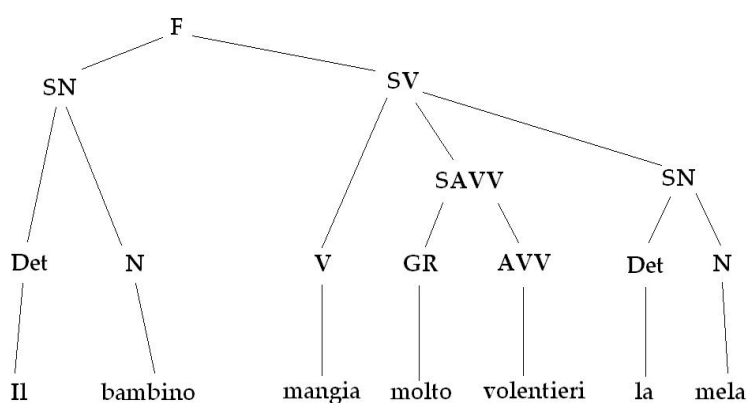
Questa sezione vuole ripercorrere brevemente i più importanti contributi scientifici della scienza cognitiva classica. È necessario sottolineare che lo scopo è quello di dare un quadro di riferimento a quello che è l'argomento principale di questa dissertazione. Pertanto questa breve storia non è da considerarsi esaustiva.

1.2.1 LA GRAMMATICA GENERATIVA DI CHOMSKY E LA MACCHINA DI TURING

Le tesi proposte dalla grammatica generativa di Chomsky e dalla Macchina

elaborata da Alan Turing costituiscono alcuni dei primi colpi inflitti al dominio epistemologico e metodologico del comportamentismo.

Chomsky (1980) sostenne che la comprensione e la produzione di enunciati linguistici è possibile grazie alla conoscenza di una grammatica universale innata e inconscia. Disporremmo, cioè, già alla nascita, di quelle strutture e regole sintattiche che ci permettono di articolare e comprendere il linguaggio. La struttura sintattica è una struttura gerarchica (“struttura dei costituenti”) che può essere visualizzata con la rappresentazione visiva degli “alberi etichettati”. Tale modello rappresentativo non costituisce, tuttavia, la reale forma degli oggetti mentali ma è un valido strumento esplicativo soprattutto per spiegare le relazioni tra le parti. Si veda la seguente figura.



Leggenda: F= frase; SN= sintagma nominale; SV= sintagma verbale; Det = determinante; N= nome; V= verbo; SAVV= sintagma avverbiale; GR= avverbio di grado; AVV= avverbio. Le posizioni in cui compaiono le lettere sono dette “nodi”; le lettere sono le “etichette di categoria grammaticale”; ciò che occupa un nodo è detto “costituente”.

Figura presa da Marraffa (2008), pag.31

Il diagramma ad albero sopra riportato rappresenta la struttura sintattica, secondo la teoria chomskyana, dell'enunciato "Il bambino mangia molto volentieri la mela".

Chomsky mette, quindi, in rilievo l'esistenza di oggetti interni dei quali la psicologia e la nascente scienza cognitiva non potevano non tener conto. La teoria della grammatica generativa, in altre parole, mostrava concretamente l'indispensabilità di far ricorso a rappresentazioni mentali, e in che modo queste ammettessero una descrizione "di alto livello" non neurofisiologica. Secondo l'autore, vi sono diversi livelli di rappresentazione di una frase. Il passaggio da un livello all'altro è dettato da regole e passaggi puramente sintattici. Fodor allargherà questo concetto, aprendo alla spiegazione di come la sintassi possa portare contenuto e quindi reggere la semantica (Marraffa 2008).

L'importanza della Macchina di Turing (di seguito MT) consiste, invece, nell'aver mostrato in che modo un processo cognitivo potrebbe essere meccanizzato: con la sua teorizzazione il concetto di computazione trova la sua formalizzazione più adatta alla riproduzione su un meccanismo e si inizia ad identificare i processi di elaborazione umani a quelli di un calcolatore artificiale.

Si immagini un nastro infinito, diviso in caselle, sul quale si sposta un cursore capace di leggere e (sovra)scrivere sulle singole caselle. Tale macchina è capace di eseguire istruzioni che specificano, dato lo stato attuale della MT e il simbolo che si legge nella casella corrente, lo stato in cui andare, la direzione nella quale spostarsi e il simbolo che bisogna scrivere nella casella prima di passare alla successiva. Ovvero l'istruzione "q1sq2s'R" significa "se sei nello stato q1 e leggi il simbolo "s", vai nello stato q2; sovrascrivi "s" e spostati a destra". In tal modo la macchina è capace di eseguire qualsiasi procedimento

effettivo. La macchina si blocca con l'istruzione finale qH. Sino a quando non trova tale istruzione, il cursore continua a spostarsi.

La MT dimostra che processi mentali anche molto sofisticati, come il ragionamento o la risoluzione di problemi, ammettono una descrizione computazionale; altrimenti detto, il processo è un'elaborazione di simboli governata da leggi. Inoltre, sebbene la MT sia un modello astratto, essa mostra nondimeno in che modo simboli e procedimento potrebbero essere realizzati fisicamente. La MT, d'altra parte, è un modello "scarnificato" di un computer reale (Marconi 2001).

1.2.2 IL FUNZIONALISMO

L'analogia tra stati logici della MT e stati mentali umani e tra gli stati dell'hardware di un computer e gli stati cerebrali ha portato a parlare dei processi mentali umani nei termini di "stati mentali" e di "impressioni" (come i simboli della macchina impressi sul nastro). Assieme all'articolo di Putnam "*Minds and Machines*" (nel quale si sosteneva che il processo mentale può essere descritto a livello logico come sequenze di stati mentali governate da regole e che una MT può avere più realizzazioni in quanto indipendente dal suo supporto materiale) e all'influenza della proposta di Chomsky si posero le basi di quella che è una delle principali teorie della scienza cognitiva: il funzionalismo.

Idea fondamentale del funzionalismo è che uno stato mentale è quel che è in virtù del ruolo che svolge nel sistema. In questo modo si prescinde dalla realizzazione fisica, che diventa ininfluente, per concentrare l'aspetto analitico attorno alla sola funzione dello stato o oggetto mentale. I processi mentali sono visti come sequenze di stati governate da leggi che ne controllano l'ordine. L'intera attività mentale è, quindi, vista come una serie di relazioni

causali e catene di stati. Tali relazioni fanno riferimento: al rapporto tra ambiente e stati interni (input percettivi); al rapporto tra stati interni e stati interni; al rapporto tra stati interni e comportamenti (output comportamentali). Gli stati mentali sono individuati nei termini del ruolo causale che svolgono. L'analisi funzionalista cerca di individuare, all'interno del sistema, i sottosistemi funzionali che fanno riferimento a dati ruoli. Teoricamente, svolgendo un percorso a ritroso dove si cerca di individuare, via via, i diversi sottosistemi, si dovrebbero individuare i sottosistemi elementari. Non a caso, molte delle rappresentazioni esplicative vedono diagrammi a blocchi (*black boxes*) che si diramano, eventualmente, in sotto-blocchi (Paternoster 2002). I fondamenti del funzionalismo non specificano l'esatto rapporto tra materia e mentale. Esistono varie proposte in merito.

Per la prima, gli stati mentali sono identici agli stati cerebrali. Ogni evento-stato-processo psicologico era ritenuto identico a qualche fenomeno neurofisiologico. Tale teoria riduzionista è nota come “teoria dell'identità di tipo” (*type identity theory*). E' necessario specificare che esiste una differenza tra tipo (*type*) e occorrenza (*token*). Una parola in un dizionario è un tipo di innumerevoli occorrenze (parlate, scritte etc). L'identità di tipo implica l'identità di occorrenza ma non viceversa. Secondo la teoria dell'identità di tipo tutte le occorrenze di un particolare tipo di stato psicologico (es. il dolore) sono identiche alle occorrenze di un tipo correlato di stato neurologico.

La teoria dell'identità di occorrenza (*token identity theory*) implica, invece, che ogni occorrenza di un evento mentale sia identica ad un evento fisico. Questo non comporta che esista un procedimento per descrivere qualsiasi tipo, o classe, di entità mentali in termini fisici. Tale identità è usata da molti filosofi per spiegare la relazione tra stati della mente e stati del cervello, in quanto è una versione più debole della teoria precedente (Marraffa 2008).

La teoria dell'identità di tipo venne messa in crisi dall'assai accreditato argomento della "realizzabilità multipla" proposto da Putnam (1960). Uno stesso tipo di stato mentale può essere realizzato da stati fisici di tipo differente che non condividono tra loro alcunché di significativo. Questa tesi si oppone fermamente alla teoria dell'identità di tipo e agli argomenti riduzionisti in generale. Fodor (1974) definì meglio la realizzabilità multipla precisando che uno stato psicologico è determinato dalle disgiunzioni di tutti gli stati fisici in cui può essere realizzato. Ovvero, lo stato psicologico del credere che la neve sia bianca è uguale alla disgiunzione di: stato fisico 1 (stato cerebrale degli esseri umani); stato fisico 2 (stato cerebrale del cane); stato fisico 3 (stato dei circuiti dell'androide)... stato fisico n (stato cerebrale del marziano). Queste sono le occorrenze fisiche della credenza "credere che la neve sia bianca" (nei limiti in cui si può attribuire a un cane la credenza che la neve sia bianca). Parlare di disgiunzione ha senso perché è sufficiente la realizzazione di almeno uno di questi per dire che si è verificata tale credenza. Parlare di equivalenza sarebbe sbagliato perché pare ovvio che l'uomo, nel nutrire una credenza sul colore della neve, non ha gli stessi stati fisici interni del cane o del marziano.

In tal senso, il sistema si trova in un certo stato psicologico *non* in virtù della sua particolare costituzione materiale *ma* in virtù dello schema astratto di relazioni causali che le parti del sistema intrattengono tra loro.

Altra spiegazione della relazione tra mentale e cerebrale è quella della "sopravvenienza". Secondo tale tesi, dato uno stato fisico ne risulta univocamente determinato lo stato mentale corrispondente. Ne segue che quando l'organismo si trova in stati cerebrali identici, in tempi differenti, anche i corrispettivi stati mentali saranno identici. La sopravvenienza è una tesi positiva: l'identità di stato fisico indica identità di stato mentale. La

soppravvenienza si basa su tre assunti: covarianza (qualunque differenza del mentale deve avere riscontro fisico), dipendenza (gli stati sopravvenienti dipendono da quelli fisici sottostanti) e irriducibilità (il mentale, tuttavia, mantiene una sua autonomia visto che un tipo di stato mentale non è riducibile). Il connubio tra funzionalismo e sopravvenienza da un lato si accorda bene col fisicalismo e dall'altro salva il mentale come livello di spiegazione autonomo (Marraffa 2008, Paternoster 2002).

1.2.3 LA TEORIA COMPUTAZIONAL-RAPPRESENTAZIONALE DELLA MENTE E LA TESI DELLA MODULARITÀ

La Teoria computazionale-rappresentazionale della mente (TCRM) è una forma di funzionalismo, in particolare è una forma computazionale di funzionalismo. Secondo questa teoria, infatti, i processi cognitivi sono computazioni condotte su rappresentazioni mentali (i cosiddetti simboli del linguaggio del pensiero, il mentalese) che condividono alcune proprietà sintattiche e semantiche con le espressioni delle lingue naturali. Le relazioni tra gli oggetti interni hanno carattere computazionale piuttosto che associativo e le operazioni che con essi vengono compiute sono guidate da un programma (regole sintattiche del linguaggio del pensiero) (Paternoster 2002).

La TCRM unisce la tesi del carattere computazionale della cognizione con quella del carattere rappresentazionale degli stati mentali. Essi sarebbero, dunque, valutabili semanticamente, ovvero dotati di condizioni di verità. Avendo natura linguistica, tali rappresentazioni presentano le seguenti caratteristiche: a) hanno parti costituenti che si combinano tra loro secondo una sintassi; b) hanno parti atomiche che si riferiscono a stati o oggetti del mondo; c) sono composizionali (ovvero dipendono dai costituenti e dalla loro combinazione sintattica); d) il loro valore di verità è legato allo stato o oggetto

del mondo cui si riferiscono; e) intrattengono tra loro relazioni logiche di implicazione (Paternoster 2002; Marraffa 2008).

Fodor arriva alla definizione di questa teoria percorrendo diverse strade. Da fautore del realismo intenzionale, secondo il quale gli assunti sugli stati intenzionali così come postulati dalla teoria psicologica del senso comune sono corretti, sposa la tesi del realismo scientifico intenzionale, secondo il quale una scienza cognitiva matura dovrà rendere conto di stati che sono valutabili semanticamente, dotati di struttura sintattica e efficaci causalmente. In particolare, la parte della rappresentazionalità viene elaborata come tesi sulla natura degli atteggiamenti proposizionali, la parte sulla computazionalità è relativa alla natura dei processi cognitivi e la parte che si dedica alla natura causal-informazionale del significato studia i fondamenti naturalistici dell'intenzionalità.

Il mentalese spiega due proprietà dei sistemi cognitivi: la produttività e la sistematicità. Un sistema così formato può, infatti, produrre pensieri infiniti a partire da un repertorio di concetti limitato. Questo grazie alla caratteristica della composizionalità (punto c) che apre alla possibilità di comporre, in modo virtualmente infinito, le varie componenti atomiche e composte.

La sistematicità consiste nel fatto che la capacità di produrre e afferrare pensieri è legata alla capacità di produrre e afferrare altri pensieri. Anche qui entra in gioco la composizionalità. Se io sono capace di capire “Mario ama la ragazza” sarò anche capace di capire “La ragazza ama Mario” (Marraffa 2008).

Ci si chiederà in che modo funziona un sistema così formulato e come è capace di produrre e trasmettere significati.