

UNIVERSITA' DEGLI STUDI DI GENOVA

FACOLTA' DI LETTERE E FILOSOFIA

CORSO DI LAUREA IN FILOSOFIA

La tematica della mente come emergenza
Da Popper alla "terza cultura"

Tesi di laurea

Relatrice

Prof. A. Papone

Correlatore

Prof. C. Penco

Candidato

Tullio Tinti

Anno accademico 1996-1997

Alla mia famiglia con amore

Indice

Introduzione	pag. 6
Parte I. Il problema mente-corpo nella filosofia contemporanea	16
1. I limiti dei paradigmi tradizionali e l'approccio personologico	
1.1 La distanza epistemologica dalla biologia e il dualismo interazionista	17
1.2 Il materialismo eliminativista: la mente è una "cattiva abitudine"	22
1.3 La prospettiva personologica e l'emergenza secondo Margolis	27
1.4 Il funzionalismo e la metafora della mente come software	31
2. L'emergentismo "quasi-sostanziale" di Karl Popper	
2.1 La teoria dei tre Mondi: fisica, mente e contenuti oggettivi	39
2.2 L'evoluzione emergente: novità, imprevedibilità e... ragnatele	43
2.3 Lo spostamento dal problema mente-corpo a quello "io-cervello"	48
2.4 Conclusioni e osservazioni critiche	55
Parte II. Sopravvenienza ed emergenza in una prospettiva fisicalistica	62
3. Il concetto di sopravvenienza dall'etica alla filosofia della mente	
3.1 Fisicalismo e riduzionismo: impossibile separarli?	63
3.2 Il monismo anomalo di Davidson	69
3.3 Sopravvenienza, covarianza, dipendenza, non-riducibilità	73
4. Il concetto di emergenza dal <i>British Emergentism</i> ad oggi	
4.1 Il primo emergentismo tra meccanicismo e vitalismo	82
4.2 L'emergenza come indeducibilità nell'opera di Broad	86
4.3 Il mentalismo emergentista di Sperry	95
4.4 Dalla causazione verso il basso al concetto di vincolo	101
4.5 Il monismo psiconeurale emergentista di Bunge	105
4.6 Conclusioni provvisorie sul concetto di emergenza	111
Parte III. Evoluzionismo ed emergentismo nella terza cultura	124
5. Il concetto di emergenza applicato all'evoluzione	
5.1 Gratuità, contingenza ed emergenza nella teoria neodarwiniana	125
5.2 Non solo caso: il modello dello "squilibrio creativo"	137
5.3 Non solo caso: i fenomeni di auto-organizzazione	146
5.4 L'emergenza della vita come interazione tra modello e osservatore	156

6. Il concetto di evoluzione applicato alla mente	
6.1 L'evoluzione filogenetica della mente	163
6.2 L'evoluzione ontogenetica della mente	173
6.3 Evoluzione della mente ed emergenza	185
Parte IV. La mente come emergenza nella terza cultura	189
7. Il concetto di emergenza applicato alla mente	
7.1 Conoscere la mente: dalla spiegazione alla comprensione	190
7.2 La nozione di «livello» e la teoria gerarchica del reale	197
7.3 La nozione di «simbolo» e la teoria computazionale della mente	209
7.4 Il risveglio dal “sogno booleano”	217
8. Il connessionismo: reti neurali e fenomeni emergenti	
8.1 Menti, cervelli e programmi	226
8.2 Tra simboli vuoti e neuroni ciechi: i subsimboli	234
8.3 La mente come emergenza secondo il gruppo PDP	244
8.4 Dal riconoscimento percettivo alla soluzione di problemi	249
8.5 Conclusioni e osservazioni critiche	255
9. L'emergenza secondo Douglas Hofstadter: metafore e modelli	
9.1 Menti, cervelli e formiche	263
9.2 Percezione di alto livello, aloni e slittamenti concettuali	270
9.3 L'architettura emergente di Copycat	275
9.4 Osservazioni conclusive	282
Conclusioni	285
Bibliografia	298

La tematica della mente come emergenza

Da Popper alla “terza cultura”

Nella seconda metà del nostro secolo si è affermata l'opinione che la vita e la coscienza sono emanazioni naturali e necessarie delle proprietà emergenti del mondo fisico. Questo cambiamento di prospettiva modifica anche le nostre credenze, il modo in cui concepiamo noi stessi e il ruolo che ci attribuiamo nell'universo.

J. Doyne Farmer [1995, 324]

Introduzione

Mente, conoscenza, conoscenza della mente

La mente è un fenomeno del quale è difficile parlare. Questa difficoltà è dovuta al fatto che non sembra possibile uscire dalla seguente situazione: qualsiasi caratterizzazione della mente si tenti, subito qualcuno si accorge che essa si basa su assunti problematici. Naturalmente, nessuno sarebbe soddisfatto della seguente definizione: la mente è quel qualcosa la cui definizione non soddisfa nessuno. C'è in tutto questo un curioso sapore di paradosso. Come mai?

Il problema è che nessuno è in grado di circoscrivere un fenomeno e dire: ecco, questo è ciò che chiamiamo «mente» - adesso studiamolo. La mente, infatti, sfugge all'indagine oggettiva e rimanda indefinitamente al soggetto conoscente. In altre parole, *non esiste un problema della mente separato dal problema della conoscenza della mente*. Questo è il punto fondamentale. Dobbiamo rassegnarci all'impossibilità di percorrere una strada "lineare" - il soggetto che conosce l'oggetto - verso la conoscenza della mente. Dovremmo invece muoverci lungo percorsi ramificati, che solo nel loro insieme possono rappresentare il cammino verso la soluzione. Questi percorsi interagiscono l'uno con l'altro e occorrono strategie per evitare i circoli viziosi.

L'ideale sarebbe affrontare il problema della *conoscenza della mente*, ma contemporaneamente porsi il problema, molto più ampio, della *conoscenza* (in quanto tale) e quello, parte di questo, della *mente*. Questi problemi si fanno avanti, minacciosamente, con domande come le seguenti: "Che cos'è la conoscenza?", "Cosa significa conoscere qualcosa?" (problema della conoscenza); "Come possiamo conoscere la mente?", "E' possibile spiegarla?", "E' possibile comprenderla?" (problema della conoscenza della mente); "Qual è la natura della mente?", "Dove si

trova?”, “Come funziona?” (problema della mente). E’ chiaro che di fronte a domande di tale profondità ci si sente smarriti. Tuttavia, attualmente vi sono segnali che fanno sperare nella percorribilità di questo cammino, nonostante esso sia difficile e irto di trappole filosofiche.

Per quanto riguarda il problema della *conoscenza* (come tale), per esempio, nel Novecento ha cominciato a svilupparsi un atteggiamento nuovo di fronte alla *circolarità*, nel passato considerata l’ostacolo maggiore a tutti i tentativi di studiare quella che possiamo rappresentare in notazione “circolare”:



La circolarità, invece di essere temuta e quindi evitata, è lentamente diventata una delle candidate al ruolo di *categoria esplicativa* della conoscenza. A partire dalla riflessione dei cibernetici sul concetto di retroazione, e soprattutto grazie agli sforzi degli epistemologi “costruttivisti”, molto lavoro è stato fatto per precisare quest’intuizione e oggi si comincia a delineare un vero e proprio *paradigma epistemologico*, caratterizzato dall’idea della “conoscenza come costruzione”.

Per quanto concerne invece il problema della *conoscenza della mente*, alcune idee sono venute alla luce proprio negli ultimi anni. Per esempio, alla domanda sul *come* possiamo cercare di conoscere la mente, una risposta interessante - oggi di dominio pubblico - è la seguente: studiando il cervello in funzione e simulandone l’attività al computer. Prima della rivoluzione tecnologica del nostro secolo, però, queste possibilità non c’erano: non solo non esistevano le simulazioni al computer, ma ciò che si sapeva del cervello, con poche eccezioni, proveniva dalle autopsie. Oggi invece è relativamente facile, mediante la tomografia ad emissione di positroni o analisi simili, esaminare i processi cerebrali ed associarli alle varie attività mentali: basta sfogliare una rivista specializzata per scoprire quali aree cerebrali si attivano, per esempio, quando leggiamo, scriviamo, ecc. Inoltre, esistono programmi di simulazione che cominciano a riprodurre (per ora in forma assai semplificata) attività mentali circa le quali, in passato, si sapeva solo dire che erano “miracolose”, come la produzione linguistica e addirittura la creatività.

Infine, per quanto riguarda il problema della *mente*, negli ultimi decenni sono state proposte nuove idee e nuove intuizioni che potrebbero essere destinate a sviluppi

eccezionali (o a fallimenti immensamente istruttivi). La tesi si occupa di alcuni aspetti di una di queste tematiche, la “mente come emergenza”.

In nessun modo sarà possibile percorrere le strade di collegamento tra i problemi che abbiamo indicato (mente, conoscenza, conoscenza della mente). Tuttavia, ed è questo quello che si dovrebbe tenere presente, *questa tesi è pensata nell’ottica di tali collegamenti*. Chiariamo questo punto.

L’impostazione generale di questo lavoro non si auto-sostiene completamente, proprio per l’infinito rimandare della mente *come oggetto di studio* alla mente *come mezzo di conoscenza* e, da questa, alla conoscenza in quanto tale. Questo rimandare non viene negato né rifiutato, bensì presupposto. Pertanto il presente lavoro viene pensato, da chi scrive, come una piccola parte della vasta ricerca che, lungo piste parallele ma collegate, procede verso la soluzione del problema della mente e della sua conoscenza, ed andrebbe pensato, da chi legge, nella medesima ottica.

La mente: in che cosa consiste l’enigma?

Una questione fondamentale, che riguarda tanto la mente quanto la conoscenza che abbiamo di essa, è: in che cosa consiste l’enigma? Cioè, *perché* abbiamo un problema della mente e *perché* è un problema la conoscenza della mente?

La domanda non è retorica: tutti noi abbiamo a che fare con la mente e ne facciamo esperienza continuamente; dunque si potrebbe presumere che essa sia qualcosa di familiare e noto, quindi aproblematico. Tuttavia, se ci fermiamo a riflettere su tutti gli altri fenomeni di cui facciamo esperienza quotidianamente, percepiamo subito una differenza: sono tutti fenomeni appartenenti alla *dimensione oggettiva* del reale, mentre la mente appartiene a quella *soggettiva*. (Ma è proprio così? Il dolore, esempio caro ai filosofi della mente, è davvero soggettivo? E i sogni, non sono forse a metà strada tra l’oggettivo e il soggettivo?)

Di fronte a questa considerazione (di per sé già problematica), è facile avvertire l’esigenza di una spiegazione di tale differenza. *Ma una spiegazione di che tipo?*

Ecco il problema, ecco in che cosa consiste l’enigma della mente: *non abbiamo idea di cosa potremmo considerare una spiegazione soddisfacente*. Sappiamo che essa

dovrebbe in qualche modo coinvolgere il cervello. Ma in che modo? Cosa potrebbe voler dire, per esempio, “conoscere la propria mente o il proprio cervello”?

Potrebbe voler dire avere un'idea generale di come funzionano, così come i meccanici hanno un'idea di come funzionano le macchine. Potrebbe voler dire avere una spiegazione completa e dettagliata del perché la gente fa tutte le cose che fa. Potrebbe voler dire avere una comprensione completa della struttura fisica del proprio cervello a tutti i livelli. Potrebbe voler dire avere lo schema completo dei collegamenti del cervello in un libro (o in una biblioteca o in un calcolatore). Potrebbe voler dire sapere con precisione che cosa succede ad ogni istante nel proprio cervello al livello neuronico, conoscere ogni scarica neuronica, ogni modificazione sinaptica, e così via. Potrebbe voler dire avere scritto un programma che supera il test di Turing. Potrebbe voler dire conoscersi in modo così perfetto che termini come subconscio e intuizione non abbiano più senso, perché tutto è chiaro e palese. Potrebbe voler dire un qualsiasi numero di altre cose [Hofstadter 1979, 755-6].

Alcuni autori hanno riproposto, relativamente a questo problema, la celebre distinzione di Dilthey tra spiegare oggettivamente (*erklären*) e comprendere soggettivamente (*verstehen*). Forse, sostengono tali autori, è relativamente facile comprendere soggettivamente la mente, ma non lo è spiegarla nei termini oggettivi che usiamo per gli altri fenomeni naturali.

Tuttavia, in un certo senso, *la nostra tesi rovescia questo punto di vista*. Possiamo infatti dire che, in una prospettiva fisicalistica (nel senso di *non-soprannaturalistica*), “spiegare” la mente è piuttosto banale: «*la mente è semplicemente ciò che fa il cervello*» [Minsky 1985, 20]. Ciò che invece è difficile, sempre dal punto di vista fisicalista, è comprendere *intuitivamente* la natura della mente (e del sé). Pensare alla mente come all'attività del cervello non ci soddisfa dal punto di vista intuitivo: *per comprendere abbiamo infatti bisogno di metafore, di immagini, di analogie*.

La metafora della mente come programma per computer ha fatto tantissimo in questo senso. E tuttavia, negli ultimi vent'anni, molti ne hanno evidenziato i limiti. E' capitato così un fatto curioso. Alcuni autori - soprattutto i connessionisti - hanno cominciato a parlare della mente come di un *fenomeno emergente*, senza però ricollegarsi esplicitamente al primo emergentismo (corrente della filosofia inglese degli anni Venti). Inoltre, la maggior parte di questi autori applica il concetto di emergenza alla mente in modo puramente “evocativo” (spesso limitandosi a scrivere in corsivo il termine), cioè senza precisare se sta proponendo una metafora alternativa a

quella computazionale, o cos'altro. Succede anche, è inutile negarlo, che una parola divenga di moda e quindi sia usata ed abusata, senza che essa abbia un contenuto realmente valido. E' questo il caso dell'emergenza? Oppure si tratta di un concetto con un referente davvero "sostanzioso"?

Il triplice scopo della tesi è: (1) precisare il significato e la portata del concetto di emergenza, in particolare se applicato alla mente; (2) presentare vari aspetti del pensiero degli autori che condividono l'idea della "mente come emergenza"; (3) esaminare in che modo i modelli emergentisti possono aiutare la riflessione filosofica a *comprendere* la natura della mente.

Il contesto della ricerca: la "terza cultura"

Possiamo suddividere la riflessione del Novecento sulla mente in tre fasi. Queste non vanno intese come rigidamente sequenziali, perché temi, problemi, idee dell'una si ritrovano nell'altra. Tuttavia, la tripartizione che proponiamo può contribuire a contestualizzare il presente lavoro.

1) Filosofia della mente "tradizionale". Durante questa prima fase, alcuni grandi pensatori, quasi tutti filosofi di professione, hanno elaborato le loro teorie sulla mente al di fuori da qualsiasi programma di ricerca interdisciplinare *su larga scala*. Husserl, Wittgenstein, Popper hanno proposto le proprie interpretazioni della mente, diversissime tra loro, e spesso sono tornati sui propri passi, hanno rivisto, distrutto, ricostruito: ma per lo più mediante l'ausilio delle proprie capacità di riflessione.

2) "Scienza cognitiva". Nella seconda metà del secolo molti scienziati hanno cominciato a studiare sistematicamente la mente. E' nata così una corrente interdisciplinare, la scienza cognitiva, che ha unito gli sforzi di filosofi e scienziati, e che ha reso disponibili ai ricercatori appartenenti ad una disciplina i risultati dei ricercatori che lavoravano in altri campi. In particolare, vengono considerate parte della scienza cognitiva le seguenti discipline: la filosofia *funzionalista* della mente (o *cognitivismo classico*), la linguistica chomskyana, l'antropologia, la neurofisiologia, l'IA convenzionale e la psicologia post-comportamentista (o *cognitivista*).

3) “Terza cultura”. Qualcosa è cambiato negli anni Ottanta. Molte voci si sono levate ad invocare il superamento del primo cognitivismo, senza per questo interrompere il dialogo interdisciplinare della scienza cognitiva. Nuove discipline hanno fatto il loro ingresso in questa scienza, mentre altre hanno perso terreno. John Brockman, facendo il punto della situazione, ha proposto di chiamare “terza cultura” - pensandola come alternativa alla cultura umanistica e a quella tecnologica - l’attività dei ricercatori che operano nei seguenti campi: la biologia evuzionistica, la genetica, la fisica, la neurofisiologia, l’IA connessionista e la psicologia post-cognitivista. Questa “terza cultura” sta producendo una nuova filosofia naturale che si interroga, tra le altre cose, sui grandi problemi della filosofia della mente e i filosofi di professione (Dennett, tanto per fare un nome) hanno cominciato a confrontarsi con la “filosofia” della terza cultura.

Il contesto del presente lavoro è quello del passaggio dalla filosofia della mente tradizionale a quella della terza cultura. E’ nella fase di passaggio, infatti, che sempre più spesso gli autori hanno cominciato ad utilizzare, ma quasi sempre senza precisarlo, il concetto di emergenza.

Prima di presentare, in questa sede introduttiva, il percorso che si seguirà nella tesi, occorre spendere ancora alcune parole sulla filosofia “orientata scientificamente” della terza cultura. Quali sono le idee fondamentali di essa?

Un’idea centrale, ad esempio, è che il mondo non è statico ed eterno, ma si evolve nel tempo. Nel XIX secolo questa verità riguardava solo il mondo biologico, mentre nel secolo successivo l’ipotesi evuzionista è diventata valida per l’universo nel suo complesso. Questa idea ha impiegato molto tempo ad affermarsi, così come è dovuto passare un secolo perché le ipotesi di Copernico fossero confermate. Si può dire che solo in questi anni ci stiamo rendendo conto di cosa significhi una realtà in evoluzione perenne. Inoltre, questa nuova filosofia considera inutile, anzi ridicola, l’ipotesi di una intelligenza superiore responsabile della bellezza e della complessità del mondo. Si può sostenere, invece, che in un contesto biologico la materia vivente si è creata e organizzata da sé a partire da principi semplici, come la selezione naturale. Credo che lo stesso si possa affermare per le leggi della fisica e la struttura del cosmo. Il terzo tema chiave è la complessità: il mondo è intessuto di una straordinaria varietà di cose e fenomeni e questa ricchezza è essenziale e non accidentale. In questo mondo complesso e autorganizzato, le proprietà delle cose hanno una natura relazionale. La nozione di proprietà assolute, come ad esempio quella di specie biologica, è diventata altrettanto obsoleta dello spazio e tempo assoluti di Newton [Smolin 1995, 21].

In altre parole, *evoluzione naturale* al posto della Creazione e *complessità* al posto del Mistero: questi i due cardini della terza cultura. E questi, necessariamente, i cardini della filosofia della mente che ha imparato a dialogare con i biologi evuzionisti, gli informatici, i neurofisiologi della terza cultura e che tuttavia non ha perso una propria sensibilità, “filosofica” appunto, a certe importanti precisazioni come: la differenza tra un uso letterale e un uso metaforico dei concetti scientifici, l’analisi dei limiti esplicativi delle teorie (anche di quelle convalidate), il riconoscimento delle semplificazioni un po’ eccessive talvolta presenti nelle speculazioni degli scienziati sperimentali.

Piano dell’opera

L’idea della mente come “emergenza” (ovvero come “fenomeno emergente”) è relativamente semplice: *la mente emerge dal cervello nel corso dell’evoluzione (filo- e ontogenetica)*.

Tuttavia, il fatto che la mente «emerga» non è banale come potrebbe sembrare. Interpretato in senso tecnico-filosofico, il concetto di emergenza sfugge a una lettura semplice e univoca. Molti filosofi, per esempio, considerano «emergenza» un sinonimo di «non-riducibilità». Poiché la maggior parte degli autori che utilizzano il concetto di emergenza condivide un approccio allo studio della mente di tipo fisicalista, cioè “non-soprannaturalistico”, la posizione di tali autori viene spesso identificata come una forma di *fisicalismo non-riduzionista* (o *anti-riduzionista*). Al fisicalismo non-riduzionista sarà dedicato un intero capitolo (capitolo 3).

Nel corso della tesi, tuttavia, *l’emergenza non verrà trattata come semplice non-riducibilità*; anzi, verrà proposta e commentata la seguente definizione:

Un fenomeno F associato a un sistema S è detto *emergente* da un osservatore O, già dotato di un modello scientifico T di S, se e solo se: studiando S congiuntamente a F, (a) O non riscontra alcun “cambiamento fisico” non previsto da T (stesse forze, stesse sostanze); però (b) O riscontra un insieme di proprietà:

- 1) *qualitativamente nuove* rispetto alle proprietà considerate in T;
- 2) *imprevedibili* in base a T;
- 3) *indipendenti* dalle singole componenti di S.

Nel presente lavoro chiameremo *paradigma* (o *filosofia*) *emergentista* (o *dell'emergenza*) o, in una parola, *emergentismo* la concezione di quanti:

- applicano alla mente e alla vita (e/o all'evoluzione) il concetto di emergenza;
- inseriscono il problema della mente nel contesto evoluzionistico;
- condividono la cosiddetta "teoria gerarchica del reale".

Naturalmente, queste tesi costituiscono la concezione *minimale* dell'emergentismo; ciascun emergentista sostiene anche altre tesi (di importanza secondaria ai fini della nostra ricerca).

Una delle idee fondamentali che hanno guidato questa tesi è la seguente: per approfondire la tematica della mente come fenomeno emergente, non basta studiare il *concetto* di emergenza, è necessario capire qual è la *filosofia* dell'emergenza. Per questa ragione, la maggior parte della ricerca sarà dedicata non all'*emergenza*, ma all'*emergentismo*. Ciò significa che presteremo la medesima attenzione a ciascuno dei tre temi-cardine della filosofia emergentista: *emergenza*, *mente*, *evoluzione*.

Di questi temi, ugualmente importanti nel contesto della terza cultura, esamineremo tutti gli accoppiamenti possibili: *emergenza ed evoluzione* (capitolo 5), *evoluzione e mente* (capitolo 6), *mente ed emergenza* (parte IV).

La struttura della tesi è la seguente. Le quattro parti corrispondono a tre diversi contesti di discorso: il contesto della parte I è la filosofia contemporanea; il contesto della parte II è, nell'ambito della filosofia contemporanea, l'approccio "naturalistico-fisicalista"; il contesto delle parti III e IV è, nell'ambito dell'approccio "naturalistico-fisicalista", la terza cultura. Questa struttura ci ha consentito la drastica selezione di autori che ha reso possibile la trattazione di un argomento altrimenti troppo vasto e articolato, come quello della mente come emergenza.

La parte I è divisa in due capitoli. Il capitolo 1 introduce il problema mente-corpo e presenta una rassegna delle sue principali soluzioni (proposte dalla filosofia contemporanea). Il capitolo 2 è dedicato a Popper: esaminando la sua opera si incontra una formulazione dell'emergentismo interessante ed estremamente personalizzata, che abbiamo chiamato "quasi-sostanziale".

La parte II è divisa in due capitoli. Il capitolo 3 è dedicato al concetto di sopravvenienza. Tale concetto esprime, da un punto di vista logico-formale, l'idea

fondamentale del fisicalismo non-riduzionista. Il capitolo 4 è dedicato al concetto di emergenza, visto in una prospettiva rigorosamente fisicalistica. Un confronto *diretto* tra i due concetti di sopravvenienza ed emergenza non è purtroppo possibile, perché gli emergentisti non hanno mai trattato il concetto di emergenza in modo logico-formale.

Anche la parte III è divisa in due capitoli. Essa affronta, nell'ambito della terza cultura, la possibilità di applicare il concetto di emergenza ai fenomeni descritti dai modelli dell'evoluzione (capitolo 5) e la possibilità di applicare il concetto di evoluzione alla mente (capitolo 6). Per poter studiare a fondo tali possibilità, sarà necessario presentare alcuni modelli dell'evoluzione, i quali - benché illustrati solo a grandi linee - allungheranno non poco il nostro cammino.

La parte IV è divisa in tre capitoli. Essa affronta, sempre nell'ambito della terza cultura, la possibilità di applicare il concetto di emergenza alla mente. Il capitolo 7 può essere considerato un capitolo di "raccordo": un paragrafo sarà dedicato al problema della conoscenza della mente; verranno introdotte le ultime due nozioni-chiave dell'emergentismo contemporaneo («livello» e «simbolo»); tratteremo inoltre la teoria gerarchica del reale; infine, un paragrafo storico-filosofico preparerà il terreno su cui ci muoveremo negli ultimi due capitoli. Il capitolo 8 è dedicato all'IA connessionista, i ricercatori della quale applicano sistematicamente il concetto di emergenza ai fenomeni mentali. Infine, il capitolo 9 è dedicato a Hofstadter, probabilmente l'emergentista più autorevole della terza cultura: è (co)autore di una serie di programmi che simulano al calcolatore le cosiddette attività mentali "subcognitive", definisce *emergente* l'architettura funzionale di tali programmi e, come i connessionisti, applica sistematicamente ai fenomeni mentali la categoria di emergenza.

Avvertenza. Per evitare ambiguità, nel corso della tesi non è *mai* stato usato il termine «emergenza» in senso ordinario. Pertanto, *ogni volta* che nel presente lavoro compare il termine «emergenza» (citazioni escluse), esso va *sempre* inteso in senso tecnico-filosofico e *mai* in senso comune. Il termine «complessità», al contrario, è stato utilizzato talvolta con il suo senso tecnico-filosofico e *altre volte con il senso ordinario*.

Parole come le seguenti: «modello», «teoria», «tesi», «ipotesi», ecc. sono state usate con relativa disinvoltura, spesso come varianti stilistiche. In ogni caso sono state adottate le convenzioni più frequenti della letteratura.

Tutte le volte in cui il contesto del discorso lo ha consentito, «cervello» e «sistema nervoso» sono stati usati al posto di «sistema nervoso centrale».

Per non appesantire la lettura, i suffissi *-istico/a* sono stati spesso sostituiti dal suffisso *-ista* (per esempio: in luogo dell'aggettivo «fiscalistico» è stata quasi sempre preferita la forma sostantivale «fiscalista»). Ma, a seconda del contesto e seguendo le convenzioni più diffuse, talvolta siamo passati ai suffissi *-ivo/a* (per esempio: scienza «cognitiva» invece di «cognitivista» o «cognitivistica»).

Parte I

Il problema mente-corpo nella filosofia contemporanea

1. I limiti dei paradigmi tradizionali e l'approccio personologico

1.1 *La distanza epistemologica dalla biologia e il dualismo interazionista*

Che cosa si intende per «mente»? Essa appare in qualche modo legata indissolubilmente alla «*capacità di pensare*» [Hebb 1980, 28], cioè all'Io pensante o alla dinamica dei pensieri. In un certo senso, questo significato di «mente» è stato “inventato” da Cartesio nella prima metà del XVII secolo [Rorty 1979, 42-4].

Il nuovo significato è nato da un vero e proprio “groviglio” di psicologia, filosofia e religione - *ego, cogitatio* e *anima* sono inseparabili nel discorso cartesiano:

Poi, esaminando con attenzione ciò che ero, e vedendo che potevo fingere, sì, di non avere nessun corpo, e che non esistesse il mondo o altro luogo dove io fossi, ma non perciò potevo fingere di non esserci io, perché, anzi, dal fatto stesso di dubitare delle altre cose seguiva nel modo più evidente e certo che io esistevo; laddove, se io avessi solamente cessato di pensare, ancorché tutto il resto di quel che avevo immaginato fosse stato veramente, non avrei avuto ragione alcuna per credere di esser mai esistito: ne conchiusi esser io una sostanza, di cui tutta l'essenza o natura consiste solo nel pensare, e che per esistere non ha bisogno di luogo alcuno, né dipende da cosa alcuna materiale. Questo che dico «io», dunque, cioè, l'anima, per cui sono quel che sono, è qualcosa d'interamente distinto dal corpo, ed è anzi tanto più facilmente conosciuto che, anche se il corpo non esistesse, non perciò cesserebbe di essere tutto ciò che è [Cartesio 1637, 83-5].

All'argomento di Cartesio si contrappose l'obiezione del suo contemporaneo Hobbes: penso, dunque sono “pensante”; ciò non significa essere una *sostanza pensante*, così come se passeggio sono “passeggiante” senza essere una *passeggiata* [Hobbes 1641, 181]. L'“obiezione della passeggiata” era destinata ad avere meno successo dell'argomentazione cartesiana; tuttavia, nel Novecento la situazione si è rovesciata e oggi è il dualismo ad essere nel mirino dei filosofi della mente.

Ma, se la mente non è una sostanza nel senso cartesiano, qual è il suo *status* ontologico? E, se non è autonoma e indipendente dal corpo come credeva Cartesio,

qual è il suo rapporto con il cervello? Queste due domande costituiscono il cosiddetto “problema mente-corpo”, probabilmente la questione centrale di tutta la filosofia della mente moderna e contemporanea.

Non è possibile presentare tutte le risposte che, dal secolo XVII in poi, sono state avanzate dai filosofi come possibili soluzioni di tale problema. In questa sede ne presenteremo alcune, ovviamente solo tra quelle più significative, e faremo corrispondere ciascuna di esse a una corrente della filosofia *contemporanea*. Il criterio che adotteremo per classificare ciascuna risposta si basa sulla *distanza epistemologica dalla (neuro)biologia* della corrispondente corrente filosofica.

Per maggiore/minore distanza epistemologica dalla biologia di una corrente non si deve tanto intendere un grado minore/maggiore della sua componente riduzionista, quanto piuttosto un minore/maggiore grado di *compatibilità* con le più recenti acquisizioni della (neuro)biologia teorica e sperimentale.

Poiché i cardini dell’argomentazione cartesiana erano due, la *cogitatio* e la *res cogitans*, vi sono almeno tre posizioni principali che vanno esaminate: quella di chi accetta l’ontologia cartesiana relativamente ad entrambi i punti (i dualisti), quella di chi la rifiuta completamente (gli eliminativisti) e quella, intermedia, di chi condivide l’impostazione di Cartesio circa uno solo dei due termini della discussione. Purtroppo molti autori, anche contemporanei, hanno cercato di definire la propria posizione filosofica solo negativamente, in contrapposizione alla filosofia cartesiana, lasciando un grande vuoto di *teorie positive della mente*. Questa debolezza è la principale responsabile del fatto che solo le posizioni estreme, il dualismo e l’eliminativismo, sono “epistemologicamente stabili”: tutte le correnti filosofiche intermedie *tendono a cadere nell’uno o nell’altro estremo*.

Ciò premesso, esaminiamo lo schema seguente:

Max lontananza epistemologica dalla (neuro)biologia

Sia <i>cogitatio</i> che <i>res cogitans</i>	Dualismo interazionista (o cartesiano)	
<i>Cogitatio</i> ma non <i>res cogitans</i>	Approccio “soggettivistico-personologico”	
	Approccio “naturalistico-fisicalista”	Funzionalismo Monismo anomalo Emergentismo
Né <i>cogitatio</i> né <i>res cogitans</i>	Materialismo eliminativista (o eliminativo)	

Min distanza epistemologica dalla (neuro)biologia

Nella prima colonna dello schema, si trova l’“impegno ontologico” di ciascuna posizione relativo ai due termini studiati da Cartesio. Tra i due estremi si trova la posizione di quanti ritengono legittimo affermare l’esistenza della mente (in quanto *cogitatio*) benché rifiutino un impegno sostanzialista: in questo caso la mente viene concepita come l’attività, o l’insieme delle funzioni, del cervello (approccio “naturalistico-fisicalista”) oppure come il soggetto, o meglio la *persona*, che ha nel cervello il proprio supporto fisico (approccio “soggettivistico-personologico”).

Si tenga presente che in questo schema non compaiono molte importanti posizioni filosofiche, tra cui: il comportamentismo non-radical (quello *radicale* è una forma di materialismo eliminativo), le teorie dell’identità mente-cervello (identità dei *tipi* e identità delle *occorrenze*) e tutte le teorie della coscienza; alcune di queste posizioni verranno comunque presentate, per lo più in contrapposizione a quelle dello schema. Si tenga anche presente che la maggior parte degli autori, soprattutto quelli della terza cultura, non è facilmente collocabile in un’unica corrente di pensiero: vi è chi cambia idea e chi “personalizza” i concetti-chiave del proprio orientamento; entrambi si sottraggono alla rigidità delle classificazioni.

Il *dualismo interazionista* è forse la posizione filosofica contemporanea fra tutte più criticata e che riscuote meno consensi. Ciò non deve stupire, perché essa sottoscrive un’ontologia - quella cartesiana - prescientifica (ovvero scientificamente premoderna) che oggi può essere sostenuta solo per coerenza con le proprie

convinzioni religiose. I limiti di questo paradigma non verranno evidenziati dettagliatamente, ma in qualsiasi manuale di filosofia della mente si possono trovare svariate critiche più o meno sistematiche. Qui ci si limiterà a tratteggiare la teoria del più autorevole dualista interazionista: il neurofisiologo John C. Eccles.

Eccles ritiene che la mente, da lui e da Popper chiamata «Mondo 2», sia un'entità primaria nettamente distinta dal mondo materiale (il «Mondo 1»), priva di massa ed energia, ma in grado di influenzare il cervello. “Prove” di tale interazione sarebbero gli esperimenti atti a misurare l'attività cerebrale associata alle diverse attività mentali. In uno di questi, per esempio, al soggetto - sdraiato al buio e in silenzio - viene chiesto di concentrarsi su un pensiero specifico, per esempio l'attesa di uno stimolo tattile ad un dito. Mentre il soggetto è in paziente attesa, 254 contatori Geiger applicati al suo cranio segnalano una forte concentrazione di sangue (in cui è stato precedentemente iniettato xeno radioattivo) nell'area sensoriale corrispondente al dito. Secondo Eccles, ciò “dimostra” che la concentrazione del soggetto, cioè un'attività puramente mentale, *causa* l'aumento del flusso ematico - attività puramente cerebrale. E, poiché le aree cerebrali attivate sono composte da gruppi di neuroni che Eccles chiama «dendroni», è *«lecito chiedersi, a questo punto, se l'attenzione mentale abbia una composizione granulare fine corrispondente alle migliaia di dendroni su cui essa agisce»* [Eccles 1991, 65].

Al di là del valore di ciò che Eccles considera una “dimostrazione” (e che per essere interpretata come tale deve presupporre ciò che va dimostrato, cioè un dualismo di sostanze causalmente interagenti), la sua risposta alla precedente domanda è affermativa: anche le esperienze mentali hanno una composizione *granulare*, così come il cervello (!). L'ipotesi di Eccles è quindi che

la totalità degli eventi e delle esperienze mentali, ovvero l'intero insieme dei sensi interni ed esterni del Mondo 2, si componga di eventi mentali unitari o elementari che potremmo definire *psiconi*. Il mondo mentale sarebbe dunque *microgranulare* e discontinuo. Si assume inoltre che ogni psicone sia legato al proprio dendrone in modo reciproco ed esclusivo. [...] Tale legame intimo fra dendrone e psicone può indurre a credere che la nuova ipotesi unitaria del dualismo interazionista altro non sia che un'elaborazione dell'ipotesi materialista dell'identità di Feigl [...]. Ma non è così. Gli psiconi hanno infatti un'esistenza indipendente e sono tutti componenti del Mondo 2 [Eccles 1991, 67].

Il rapporto tra psiconi e vita mentale è molto chiaro: gli psiconi *sono* le esperienze mentali stesse [Eccles 1991, 67].

Sul come funzionino effettivamente gli psiconi, Eccles sembra avere le idee meno chiare e si limita ad un'analogia: uno psicone agirebbe *analogamente* ad un *campo di probabilità quantistica* e la sua attività si limiterebbe ad influenzare il comportamento delle particelle coinvolte nella trasmissione sinaptica dei neurotrasmettitori [Eccles 1991, 69].

Si consideri di nuovo l'esperimento sul soggetto concentrato, in attesa di uno stimolo tattile, i cui pensieri (in questo caso l'attenzione) pre-attivano l'area tattile della corteccia cerebrale. Supponiamo ora che il soggetto venga effettivamente stimolato; ecco quello che accadrebbe secondo Eccles: lo stimolo percettivo attiva il «*dendrone connesso allo psicone che presiede alla percezione tattile*» [Eccles 1991, 70]; a questo punto lo psicone si trova di fronte ad un aumento di particelle pre-sinaptiche pronte a trasmettere i neurotrasmettitori attraverso le sinapsi e su di esse agisce selettivamente producendo «*direttamente l'esperienza della percezione tattile, la quale sarà integrata nel Mondo 2*» [Eccles 1991, 71].

La soluzione al problema mente-corpo proposta da Eccles non è più "scientifica" di quella di Cartesio solo perché sfrutta i concetti della fisica quantistica: si tratta di uno sfruttamento che utilizza parole e idee fuori luogo con lo scopo di evocare in chi legge la sensazione di chissà quali sofisticati «*calcoli basati sul principio di indeterminazione di Heisenberg*» [Eccles 1991, 68]; in altre parole Eccles vuol dare una veste scientifica a quella che è, e rimane, la posizione di un uomo di fede. Egli stesso ammette: «*ritengo essenziale postulare la creazione sovranaturale di ciascun Io cosciente quale mistero che sfugge alla comprensione scientifica*» [Eccles 1991, 73-4].

La teoria dei due Mondi di Eccles fa parte di una teoria di più ampio respiro dovuta al famoso epistemologo Karl Popper. Prima di esaminare quest'ultima è però necessario illustrare brevemente altre posizioni filosofiche, a cominciare da quella diametralmente opposta al dualismo interazionista: il materialismo eliminativista.

1.2 Il materialismo eliminativista: la mente è una “cattiva abitudine”

Parlare della mente potrebbe essere solo una “cattiva abitudine”? La domanda sembra a prima vista del tutto insensata, ma la storia del pensiero insegna che molte volte, in passato, è sembrato “insensato” mettere in discussione l’esistenza di enti che oggi vengono universalmente considerati fittizi. Anche la ricerca scientifica ha inseguito “chimere” di questo tipo: il calorico, il flogisto, i flussi vitali e l’etere ne sono esempi ben noti. Tenendo conto di ciò, alcuni filosofi hanno cominciato a prendere in considerazione la possibilità che la mente possa essere una di tali “chimere”, cioè che in realtà *non esista*. Ma allora, ci si deve chiedere, di *che cosa* parla la gente quando parla di «fenomeni mentali»?

Una risposta potrebbe essere: in realtà, senza saperlo, ci riferiamo a processi cerebrali ogni volta che parliamo di pensieri o stati mentali. E’ solo perché la neurobiologia è una scienza ancora “immatura” che il senso comune parla di «pensieri», «dolore», «immagini mentali»: quando le neuroscienze saranno più avanzate, sapremo a quali stati cerebrali *corrispondono esattamente* tali fenomeni mentali e, volendo, potremo parlare in termini neurologici anziché mentalistici. Quest’idea è nota in filosofia della mente come “teoria dell’identità”. Secondo la formulazione originaria di questa teoria, sostenuta tra gli anni Cinquanta e Sessanta da Herbert Feigl e dai filosofi della cosiddetta “scuola australiana” (Place, Smart e Armstrong), e oggi nota come “identità dei tipi”, *gli stati mentali sono (identici agli) stati cerebrali*. Tale identità dovrebbe comportare una strettissima correlazione tra determinati tipi di eventi mentali e determinati tipi di eventi cerebrali, correlazione che però non sembra affatto esistere. Nessun genere di ricerca neurobiologica ha mai mostrato, per esempio, un tipo specifico di processi cerebrali associati al “credere qualcosa”, oppure allo “sperare qualcosa”, ecc.

Il punto di partenza dell’*eliminativismo* è proprio la constatazione che le neuroscienze non sono in grado di dimostrare alcuna corrispondenza, del genere di quelle previste dalla teoria dell’identità dei tipi, tra mente e cervello. Evidentemente, sostengono i materialisti eliminativi, se non vi è identità, allora “cose” come i fenomeni mentali sono destinate a rivelarsi inesistenti. Prima di illustrare la critica che si può rivolgere a questa inferenza eliminativista, critica che si basa sulla convinzione

che *la mente andrebbe spiegata* - o, meglio ancora, *compresa - e non eliminata*, occorre fare un'importante distinzione tra due versioni dell'eliminativismo leggermente differenti, anche se *non* mutuamente esclusive (si tratta di una differenza resa spesso invisibile dal fatto che molti autori adottano ora l'una, ora l'altra versione, anche all'interno della medesima argomentazione).

Consideriamo dunque l'ipotesi eliminativista secondo cui *non esiste una cosa come la mente*. E' possibile far propria quest'ipotesi da due punti di vista differenti: quello, più radicale, di chi nega l'*esistenza della mente* e quello, relativamente meno radicale, di chi nega la *veridicità del linguaggio* che ne parla (la cosiddetta "psicologia del senso comune"). Chiamiamo rispettivamente queste due prospettive: *eliminativismo verso la mente* ed *eliminativismo verso la psicologia del senso comune*. Queste concezioni possono differenziarsi a loro volta - ciascuna in una forma più "forte" e in una più "debole". L'eliminativismo verso la mente comprende infatti la posizione di chi (1a) è eliminativista nei confronti del referente *in quanto tale* di «mente» (per quanto diverso esso possa essere da come il senso comune se lo immagina) e quella, più debole, di chi (1b) è eliminativista nei confronti della mente *in quanto sostanza*; l'eliminativismo verso la psicologia del senso comune comprende invece la posizione di chi (2a) ritiene falsa, e destinata ad essere soppiantata, la psicologia del senso comune e quella, più debole, di chi (2b) la ritiene falsa, e quindi eliminabile in teoria - *ma non in pratica*. Esaminiamo con ordine queste accezioni del materialismo eliminativo.

1a) *Eliminativismo "forte" verso la mente*. La prima forma di materialismo eliminativista, la più radicale fra tutte, sostiene che non esiste alcunché - né sostanze, né eventi, né proprietà - a cui ci si riferisce parlando di fenomeni mentali. Una teoria di questo tipo venne difesa dai *comportamentisti radicali* (e in particolare da John B. Watson) a partire dal 1913. In base ai risultati della neurofisiologia del loro tempo, i comportamentisti radicali supposero che il comportamento non avesse cause mentali e trattarono idee e immagini mentali come ingannevoli abitudini verbali. Oggi sappiamo che i dati di cui disponevano questi studiosi erano gravemente incompleti [Hebb 1980, 37]. Chi scrive ritiene però che, tuttora, questa forma di eliminativismo venga spesso imputata, *erroneamente*, a quanti ne sostengono la versione debole: l'eliminativismo verso la mente come sostanza (1b).

1b) *Eliminativismo “debole” verso la mente*. Gli autori che negano l’esistenza di *sostanze* mentali sono molti di più rispetto a quanti si dichiarano materialisti eliminativi: in pratica, la maggioranza delle correnti filosofiche contemporanee ammette l’esistenza di processi mentali (quindi della mente-*cogitatio*) negando risolutamente l’esistenza di *cose* come i fenomeni mentali (quindi della mente-*res cogitans*). Questa prospettiva è quella più naturale per quanti ricordano nelle loro pagine, con preoccupazione, gli “abbagli” storici legati al flogisto, all’etere, ecc. [cfr. Bellone 1992, 11]: in effetti un simile abbaglio potrebbero aver preso anche Cartesio, che parlava della mente come di una “cosa”, ed Eccles, che la considera un’entità primaria ontologicamente autonoma.

Il filosofo della mente Daniel C. Dennett, i cui epiteti sono sempre molto efficaci, ha impostato il problema parlando di oggetti teoretici “buoni” e “non-buoni”. Enti teoretici buoni sono quelli tipici del linguaggio non-mentalistico, come gli elettroni e le cellule nervose; oggetti teoretici non-buoni sono le entità mentali putative tipo: il credere qualcosa, il dolore, le immagini mentali. Per chiarire il suo concetto di oggetto teoretico non-buono, Dennett ricorre ad un esperimento mentale:

Supponiamo di scoprire una società che non ha la nostra conoscenza della fisiologia umana, e che usa una lingua del tutto simile alla nostra, tranne che per un curioso gruppo di idiotismi. Quando sono affaticati, dicono di essere assaliti da *stanchezze*, di avere *stanchezze* mentali, *stanchezze* muscolari, *stanchezze* negli occhi e *stanchezze* dello spirito. La loro tradizione sportiva contiene massime del tipo: «Troppe *stanchezze* danneggiano il fine» e «Cinque *stanchezze* nelle gambe valgono quanto dieci nelle braccia» [Dennett 1978a, 30].

Le entità putative che questa società immaginaria chiama *stanchezze* sono entità teoretiche non-buone. Ora, secondo Dennett, si dovrebbe assumere una posizione eliminativista nei confronti degli oggetti teoretici non-buoni presenti nella psicologia del senso comune *solo se si è certi che tali oggetti non potrebbero, neppure in una psicologia matura, essere sostituiti da entità teoriche buone* [Dennett 1978a, 31].

Probabilmente, se si sostiene questa posizione, si dovrebbe usare *molta* cautela nel dichiararsi materialisti eliminativi, perché si rischia di essere scambiati per eliminativisti “forti” verso la mente, quando invece si sta sostenendo una posizione accettata dalla maggioranza dei filosofi della mente.

2a) *Eliminativismo “forte” verso la psicologia del senso comune.* La terza forma del materialismo eliminativo è sicuramente quella più caratteristica. Formulata nei primi anni Sessanta dagli epistemologi Paul K. Feyerabend e Richard Rorty (che ha poi differenziato la propria posizione da quella di Feyerabend), essa ha continuato a riscuotere sporadici consensi. Ecco come il filosofo della mente Paul M. Churchland descrive questa posizione:

Il materialismo eliminativo è la tesi secondo la quale la concezione che comunemente abbiamo dei fenomeni psicologici costituisce una teoria radicalmente falsa - una teoria così manchevole che sia i suoi principi sia la sua ontologia finiranno per essere soppiantati, invece di essere progressivamente ridotti, una volta che le neuroscienze saranno state completamente sviluppate. La nostra reciproca comprensione e perfino la nostra introspezione potranno allora esser ricostituite all'interno della cornice concettuale delle neuroscienze, ovvero di una teoria che, secondo quanto possiamo attenderci, sarà di gran lunga più potente della psicologia del senso comune che verrà a sostituire e, in generale, più sostanzialmente integrata con la scienza fisica [Churchland 1981, 29].

Il nucleo della psicologia del senso comune, così come viene attaccata da Churchland, è costituito da quelli che Bertrand Russell chiamava “atteggiamenti proposizionali”. Esempi di atteggiamenti proposizionali sono: *credere che...* “qualcosa”, *sperare che...* “qualcosa”, *sapere che...* “qualcosa”, ecc., dove «qualcosa» può essere sostituito da una proposizione. La psicologia del senso comune spiega e prevede (più o meno efficacemente) il comportamento delle persone facendo affidamento sulla propria concezione della *mente come insieme, sede o successione degli atteggiamenti proposizionali degli individui.*

Churchland insiste molto sul fatto che la psicologia del senso comune (cioè l'apparato di ascrizione di atteggiamenti proposizionali usato per descrivere stati mentali) vada considerata una vera e propria *teoria empirica* e che si debba prendere sul serio la possibilità che essa, come teoria empirica, sia *radicalmente falsa*, né più né meno di quanto lo era — per esempio — la teoria alchimistica della materia inanimata¹.

¹La ragione di questo fatto è da individuarsi in tre difficoltà in cui incappa la psicologia del senso comune: la sua scarsa integrazione nella neuroscienza, l'esistenza di fatti inspiegabili in termini di atteggiamenti proposizionali e il suo “ristagno” durato millenni, senza miglioramenti degni di nota. Relativamente alla prima difficoltà, si è già detto che essa costituisce il punto di partenza di tutta la riflessione eliminativista; per quanto concerne l'esistenza di fatti che restano misteriosi all'interno della psicologia del senso comune, Churchland ha in mente la natura e la dinamica delle malattie mentali, la facoltà dell'immaginazione creativa, il sonno, la visione tridimensionale, le illusioni percettive e il

2b) *Eliminativismo “debole” verso la psicologia del senso comune*. La quarta versione del materialismo eliminativista può essere considerata una versione “*soft*” della posizione di Churchland. Questa versione è così debole che non impone quello che il filosofo William Bechtel chiama l’*aut-aut* dell’eliminativismo: o prospettiva mentalista o prospettiva neuroscientifica, *ma non entrambe* [Bechtel 1988, 174].

Si tratta della tesi secondo cui:

espressioni come «credenza», «desiderio», «paura», «speranza», ecc. andrebbero intese come semplici espedienti verbali, privi di qualunque riferimento a fenomeni mentali intrinseci. Una terminologia siffatta può essere utile per spiegare il comportamento e formulare delle previsioni sul medesimo, ma non va presa alla lettera come se ci si riferisse a fenomeni mentali reali, intrinseci e soggettivi. Coloro che aderiscono a questo punto di vista ritengono che la scelta di utilizzare la terminologia del senso comune dipenda dall’«atteggiamento intenzionale» che si assume nei confronti di un certo sistema [Searle 1992, 23].

Il già ricordato Daniel Dennett, che chiama “atteggiamento intenzionale” l’atteggiamento di chiunque usi la psicologia del senso comune per prevedere il comportamento di qualcosa (persona, animale, computer), è certamente un rappresentante di questa posizione. Egli precisa: *la «decisione di adottare questa strategia è pragmatica, non è intrinsecamente giusta o sbagliata»* [Dennett 1971, 44].

Una delle critiche più efficaci alle forme “forti” dell’eliminativismo, ma in particolare a quella di Churchland, è dovuta al filosofo Andy Clark. Egli pone al centro dell’attenzione il problema della *spiegazione* e osserva giustamente che eliminare le spiegazioni della psicologia del senso comune ci priverebbe di «*spiegazioni legittime e perfettamente intelligibili sia nell’ambito della scienza cognitiva che in quello della vita di tutti i giorni*» [Clark 1989, 279]. La ragione di ciò è che non sempre, tra due forme di spiegazione, quella più precisa (nel nostro caso quella neurologica) è più *potente* dell’altra (nel nostro caso quella del senso comune).

“miracolo” della memoria [Churchland 1981, 36]; infine, la terza critica riguarda il fatto che la psicologia del senso comune usata nell’antica Grecia, due o tremila anni fa, era sostanzialmente quella che usiamo oggi: è «*questo un periodo di stagnazione e sterilità eccessivamente lungo per qualunque teoria [...]. Nei termini di Imre Lakatos, la psicologia del senso comune è un programma di ricerca stagnante o in degenerazione*» [Churchland 1981, 38]. Per tutte queste ragioni, Churchland conclude che la psicologia del senso comune potrebbe essere *realmente* falsa e irriducibile al linguaggio delle neuroscienze “mature”.

Per chiarire l'idea, Clark ricorre al seguente esempio: consideriamo un direttore d'orchestra infastidito dal fatto che qualcuno (Fred) ha tossito durante il concerto e

supponiamo che qualcuno, nell'interesse dell'«accuratezza», insista che la spiegazione appropriata (completamente causale) del senso di fastidio del direttore d'orchestra sia in realtà il tossire *di Fred*. Vi è un senso buono in cui la loro spiegazione «più accurata» è in realtà *meno potente*. Infatti la spiegazione che usa il termine «qualcuno» ha il vantaggio di chiarire che «chiunque dei membri dell'uditorio avesse tossito, avrebbe causato un disturbo al direttore d'orchestra» [...]. Questo aumento in generalità, ottenuto sacrificando l'indicazione della reale identità implicata nella catena causale in questione, costituisce una virtù esplicativa che legittima tutta una serie di spiegazioni causali [Clark 1989, 266].

Applicando questa osservazione al problema della mente, è lecito affermare, seguendo Clark, che *l'efficacia esplicativa della psicologia del senso comune è una questione del tutto indipendente dalla sua veridicità*. Ciò significa che, anche se la psicologia del senso comune fosse radicalmente falsa, come sostiene Churchland, non per questo essa dovrebbe essere considerata una “cattiva abitudine”, cioè una teoria non-esplicativa ed eliminabile.

1.3 La prospettiva personologica e l'emergenza secondo Margolis

Dopo aver tratteggiato le soluzioni estreme al problema mente-corpo, il dualismo interazionista e il materialismo eliminativista, dobbiamo cominciare ad esaminare la posizione di quanti, e sono la grande maggioranza, rifiutano entrambe per percorrere una sorta di “terza via”.

Per come affrontano il problema mente-corpo, i filosofi e gli scienziati appartenenti a questa vasta “terza via” possono essere divisi in due raggruppamenti principali, corrispondenti alle due diverse immagini dell'essere umano che, secondo lo psicologo sociale John Shotter, si vanno definendo e contrapponendo nel pensiero contemporaneo. Una è l'immagine dell'uomo-organismo, completamente determinato dalle proprietà fisiche del suo corpo; l'altra è quella dell'uomo-persona, «*produttore di atti, simboli e valori connessi essenzialmente con la sua natura storico-culturale*» [Moravia 1986, XII]. La distanza tra queste due visioni dell'essere umano è enorme.

Relativamente al problema della natura della mente, esse cercano la soluzione in direzioni opposte: verso lo studio del funzionamento del cervello, la prima; verso l'analisi della dimensione della soggettività, la seconda. Poiché tutti gli autori (Popper escluso) di cui ci occuperemo nel resto della tesi appartengono al primo gruppo, il cui approccio chiameremo "naturalistico-fisicalista", per ragioni di completezza dedicheremo questo paragrafo all'*altro* gruppo, la cui importanza non può essere sottostimata, benché si possa non dividerne la prospettiva.

Innanzitutto, si osservi che in questo gruppo di pensatori confluiscono diverse e importantissime tradizioni filosofiche, prima tra tutte quella *fenomenologica*. Secondo i fenomenologi,

di fronte alla difficoltà di individuare quel paradossale *oggetto* che è la mente, un'entità pubblicamente riconoscibile da collocare, sia pure con le specificità del caso, nell'ambito dell'ordine naturale, esistono due soluzioni: quella (sbagliata) eliminativistica, che dissolve l'*illusione* dell'io, mostrandone la presunta inconsistenza teoretica e quella (corretta) che interpreta la presenza ineliminabile dell'io non nei termini di un dualismo sostanzialista cartesiano, ma come riconoscimento di un "modo originario" in cui l'essere è dato "dall'interno" [Di Francesco 1996, 193].

Ma non sono solo i fenomenologi a parlare della mente come "modo" dell'esperienza soggettiva. Il filosofo Joseph Margolis, per esempio, suggerisce di studiare la consapevolezza individuale-soggettiva come un *modo d'essere* dell'umano [Moravia 1986, 203]. Margolis fa parte in effetti di quell'insieme di autori vicini alla fenomenologia, ma con formazioni intellettuali eterogenee, che condividono con i fenomenologi l'immagine dell'essere umano come soggetto-persona.

Il filosofo Sergio Moravia ha cercato di trovare un filo rosso (non necessariamente cronologico) che leghi tutti gli autori appartenenti a tale orientamento "soggettivistico-personologico": partendo da Rorty, che - nella seconda metà degli anni Settanta - ha abbandonato la precedente posizione eliminativista, Moravia ha messo in luce la vicinanza tra studiosi come Malcolm, Wilkerson, Nagel, Grene e Dreyfus, fino alla posizione (particolarmente interessante per noi) di Margolis. Tutti questi autori condividono una ferma condanna del dualismo e *contemporaneamente* una più o meno spiccata avversione per il materialismo. Il filosofo wittgensteiniano Norman Malcolm, per esempio, elogia la filosofia contemporanea che ha *giustamente*

voltato le spalle alla concezione cartesiana ma, di fronte alla possibilità che un *cervello* possa *avere* pensieri, illusioni o dolori, osserva:

L'insensatezza della supposizione pare così ovvia che trovo difficile prenderla sul serio. Nessun esperimento potrebbe stabilire questi risultati per un *cervello*. Perché no? La ragione fondamentale è che *il cervello* non assomiglia abbastanza a un *essere umano*.²

La differenza tra un cervello e un essere umano sta nel fatto che quest'ultimo può venire descritto, a differenza del cervello, parlando delle sue capacità correlate all'uso *responsabile* dei sistemi simbolici istituiti dall'umanità stessa, nella scienza come nell'arte, in teoria come in pratica. Nel delineare in questo modo la figura dell'essere umano, la filosofa heideggeriana Marjorie Grene definisce la *persona* un "essere-nel-mondo" [Moravia 1986, 259] e il filosofo Hubert Dreyfus, autore famoso per il suo impegno contro il meccanicismo in filosofia della mente, accoglie tale definizione:

A una possibile domanda sul "dove" avvengono gli eventi mentali la risposta più appropriata è per Dreyfus che «avvengono in un *mondo* condiviso [con altri] nel quale siamo circondati da cose e da individui esterni a noi, e *non* nei nostri *cervelli* né nelle nostre *menti*» [...]. E alla possibile domanda su "chi" sia l'agente, il soggetto reale degli atti mentali la risposta più appropriata pare essere *la persona*: non certo *la mente* [Moravia 1986, 264-5].

Ma l'autore che, tra quelli citati, risulta più interessante ai fini del presente lavoro è certamente Joseph Margolis. Margolis occupa infatti una posizione a sé stante nell'ambito dell'approccio "soggettivistico-personologico", particolarmente vicina al paradigma emergentista della mente, posizione che egli stesso colloca a metà strada tra il riduzionismo fisicalistico e il dualismo cartesiano. In effetti, il concetto stesso di *emergenza* ha un ruolo chiave nel pensiero di questo autore: la *persona* viene definita, infatti, un «ente *emergente* culturale»³.

Che cosa significa per Margolis che la persona è un «ente emergente»? Significa che la persona è un complesso di funzioni *irriducibili* alle strutture fisiologiche *ma incarnate in esse* [Moravia 1986, 269]. Il concetto di emergenza, cioè *irriducibilità* più *incarnazione* (*embodiment*), allude ad una relazione tra due

²Malcolm N. [1965], "Scientific materialism and the identity theory", in Borst C. V. (a cura di), *The mind/brain identity theory*, Macmillan, Londra 1970, pp.179-80; citato in Moravia [1986, 243].

³Margolis J. [1978], *Persons and minds*, Reidel, Dordrecht 1978, p.7; citato in Moravia [1986, 269].

dimensioni distinte, ma strettamente collegate. Cerchiamo di chiarire questo concetto tramite alcuni esempi.

Gli esempi portati da Margolis sono: la statua (il *David* di Michelangelo), la poesia, il sogno. Di che cosa è fatto il *David*? Certamente di marmo, ma non è *solo* un blocco di marmo:

Il “di più” (il “diverso”) rispetto alla materialità del blocco di marmo è l’idea, il progetto, la realizzazione specificamente “figurale” (intenzionale, semantica) dello scultore - con tutti i presupposti e le implicazioni *culturali* di ciò [...]. Tutto questo è quell’“emergente” - irriducibile alle componenti (*materiali*) dell’opera e tuttavia in esse “incarnato” - che sta tanto a cuore a Margolis [Moravia 1986, 270].

Per la poesia vale lo stesso discorso: oltre alla realizzazione fisica, vi sono in essa delle proprietà *emergenti* irriducibili: il significato, lo “spirito simbolico”, le “qualità emozionali”, la finalità e il “disegno” [Moravia 1986, 271]. Per quanto riguarda il problema mente-corpo, l’esempio utilizzato da Margolis per illustrare il proprio punto di vista è il sogno.

Nel caso del sogno, siamo di fronte ad un evento mentale indubbiamente veicolato da un sistema fisico, il cervello. Tuttavia, in nessun modo il sogno può essere considerato *coincidente* con il sistema fisico stesso:

V’è una “parte” di esso la quale, pur *incarnandosi* in certi meccanismi neurocerebrali, non è tali meccanismi. Se durante il sonno io ho un incubo, ciò di cui ho paura è non già un certo evento *fisico* bensì certe figure e vicende appartenenti ad un *altro* ordine di realtà [...]. Nessun esame pur “perfetto” dei processi che si verificano nel mio cervello mi consentirebbe di afferrare “tutto” il sogno. Per coglierne i contenuti simbolico-affettivi io devo avvalermi di *altre* indagini, di *altri* strumenti [Moravia 1986, 272].

Va qui commentata, in conclusione, la differenza tra l’emergentismo “psico-personologico” di Margolis e l’emergentismo della maggioranza degli autori che verranno incontrati nel seguito. Tale differenza è stata ben colta da Moravia: l’emergentismo facente parte di quello che abbiamo chiamato approccio “naturalistico-fisicalista” sottolinea la fondamentale unità materialistica del reale, al punto da difendere in qualche modo la teoria dell’identità mente-cervello [cfr. Bunge 1980]; Margolis al contrario sottolinea la pluralità dei piani del reale e la loro reciproca *differenza* [Moravia 1986, 269], arrivando a criticare uno dei fondatori

dell'approccio personologico (Peter F. Strawson), accusato di non evidenziare abbastanza la "consistenza" della persona [Moravia 1986, 272]. In altre parole, Margolis sembra

avvertire l'esigenza di delineare una figura di persona nello stesso tempo più "sostanziosa" e più indipendente. Di qui, tra l'altro, certi accenti "ontologizzanti" che possono creare qualche disagio [Moravia 1986, 272-3].

In questa sede non è possibile dedicare altro spazio alla prospettiva "soggettivistico-personologica". Benché essa offra ricche e interessanti analisi della soggettività umana, la sua lontananza epistemologica dalla biologia la esclude, di fatto, da qualsiasi dibattito *orientato scientificamente* sul problema mente-corpo. E' vero che essa non può essere confusa con il dualismo cartesiano, che viene ripetutamente ed esplicitamente rifiutato (il che non è poco), tuttavia tale prospettiva risente pesantemente di una certa acredine verso la ricerca scientifica, che la porta inevitabilmente a sottovalutare i risultati provenienti dalla neurobiologia.

La corretta valutazione di tali risultati è stata invece estremamente proficua per la filosofia della mente che ha saputo dialogare con gli scienziati della terza cultura.

1.4 Il funzionalismo e la metafora della mente come software

Veniamo ora, sempre relativamente al problema mente-corpo, all'approccio "naturalistico-fisicalista", il quale, nell'ambito della "terza via" compresa tra dualismo ed eliminativismo, affronta la questione cercando di accorciare la distanza epistemologica che separa la filosofia dalla biologia.

In questo campo il discorso si fa più difficile: i rapporti tra i vari indirizzi sono più complicati, il contenuto delle teorie proposte è più ostico, la quantità di letteratura prodotta è immensa. Una possibilità di districarsi in questa selva è quella di distinguere periodi e correnti in modo più rigido rispetto a come stanno realmente le cose, possibilità che utilizzeremo qui con la raccomandazione di non dimenticare gli angusti limiti di ogni classificazione teorica.

Distinguiamo allora tre paradigmi filosofici principali: *funzionalismo*, *emergentismo*, *monismo anomalo*. Il funzionalismo è il paradigma filosofico dominante della scienza cognitiva⁴. In un certo senso, recentemente anche l'emergentismo ha cominciato a reclamare un proprio spazio nella scienza cognitiva, in relazione ad una “mutazione genetica” incorsa nell'IA (la disciplina che studia l'Intelligenza Artificiale, anch'essa parte della scienza cognitiva). La mutazione è quella che ha portato - nel giro di un ventennio - da programmi per calcolatore come lo SHRDLU di Terry A. Winograd, alle reti neurali dei connessionisti. Tuttavia, per non ingenerare confusione, *ascriveremo l'emergentismo alla terza cultura e non alla scienza cognitiva*.

Poiché gli autori si riferiscono a tutti questi indirizzi utilizzando molti nomi diversi, forse può servire la seguente tabella che ne raggruppa alcuni (le date sono puramente indicative⁵):

	<i>Scienza cognitiva (1967-1986)</i>	<i>Terza cultura (1986-1996)</i>
Filosofia della mente	Funzionalismo o cognitivismo classico	Emergentismo o microfunzionalismo
IA	IA convenzionale o simbolica o BAIA (Buona Antica IA)	IA connessionista o subsimbolica o connessionismo

E' fondamentale tener presente che il funzionalismo ha avuto, *ed ha ancora*, un ruolo di primissimo piano in tutta la filosofia della mente e nell'immaginario popolare, e la sua importanza non sarà certamente messa in discussione nel presente lavoro. A differenza degli altri paradigmi della mente incontrati finora, infatti, il funzionalismo non solo ha proposto un modello della *natura* della mente, ma anche un modello della

⁴Il funzionalismo come orientamento *filosofico* (appartenente alla scienza cognitiva) non va confuso con l'omonima scuola *psicologica*. Il funzionalismo psicologico fu sviluppato dal grande psicologo William James e dal pragmatista John Dewey alla fine del secolo scorso. Tali autori sostennero, in polemica con lo strutturalismo di Edward B. Titchener, che la vita psichica umana è caratterizzata da un flusso ininterrotto di esperienze mentali e che la mente è un insieme di funzioni biologiche adattatesi nel corso dell'evoluzione. (Si noti che la prospettiva evolucionistica, presente nel funzionalismo psicologico, non ha svolto un ruolo rilevante nel funzionalismo filosofico.)

⁵Del 1967 sono sia il saggio “Psychological predicates” di Hilary Putnam, sia *Cognitive Psychology* di Ulrich Neisser. *Parallel Distributed Processing*, il testo fondamentale sul connessionismo curato da David E. Rumelhart e James L. McClelland, è stato pubblicato nel 1986. (L'IA, cioè la disciplina che studia l'Intelligenza Artificiale, divenne ufficialmente programma di ricerca già nel 1956.)

struttura della mente: la “mente modulare” [Fodor 1983]. Superando tutti i paradigmi tradizionali, inoltre, esso ha prodotto la *metafora* della mente più efficace del secolo: la mente come *software*.

La soluzione del funzionalismo al problema della natura della mente è relativamente semplice: *i fenomeni mentali sono stati funzionali*. Per cogliere appieno il senso di questa proposta, occorre confrontare quest’idea con altre due tesi in opposizione alle quali si è venuta definendo: la tesi dei teorici dell’identità dei tipi, secondo cui *i fenomeni mentali sono (identici a) stati cerebrali*, e quella dei comportamentisti, secondo cui *i fenomeni mentali sono disposizioni comportamentali*.

Il punto di partenza della riflessione funzionalistica è, come già per il materialismo eliminativo, la critica alla teoria dell’identità dei tipi. Prendiamo per esempio il fenomeno mentale del *dolore*: il teorico dell’identità sostiene che si tratta di uno stato fisico-chimico del cervello. Ma si consideri, nelle parole del filosofo Hilary Putnam, quale insostenibile compito spetta a chi vuole rendere accettabile tale tesi:

Egli deve specificare uno stato fisico-chimico tale che *qualsiasi* organismo (non solo un mammifero) prova dolore se e solo se (a) possiede un cervello con una struttura fisico-chimica opportuna, e (b) il suo cervello si trova in tale stato fisico-chimico. Ciò significa che lo stato fisico-chimico in questione deve essere un possibile stato del cervello di un mammifero, del cervello di un rettile, del cervello di un mollusco (i polipi sono molluschi, e sicuramente sentono il dolore), e così via. Nello stesso tempo, esso *non* deve essere uno stato del cervello possibile (fisicamente possibile) per nessuna creatura fisicamente possibile che non sia in grado di sentire il dolore. Anche se si riuscisse a trovare uno stato del genere, deve essere nomologicamente certo che esso sarà altresì lo stato cerebrale di qualsiasi forma di vita extraterrestre individuabile la quale sia capace di sentire dolore, prima che noi possiamo anche solo supporre che si tratti davvero di dolore [Putnam 1967, 468-9].

E lo stesso vale naturalmente per ogni altro stato psicologico: fame, rabbia, eccitazione, ecc. Poiché noi identifichiamo tali stati psicologici in base al *comportamento* degli organismi, si potrebbe pensare che le cose vadano meglio sostenendo, con i comportamentisti, che i fenomeni mentali sono disposizioni comportamentali. Ma non è così.

Il comportamentismo ha il difetto, rilevato dai funzionalisti ma non solo da loro, di ridurre ogni processo psicologico, per quanto complesso, a *relazioni tra stimoli e risposte*. Un esempio non psicologico, dovuto al maggiore esponente del cognitivismo, Jerry A. Fodor, può mostrare con semplicità la differenza tra

comportamentismo e funzionalismo, e contemporaneamente può illustrare due concetti chiave del funzionalismo: l'*interdefinizione* tra stati funzionali e la loro *indipendenza dall'hardware*. Consideriamo due macchine, una comportamentista ed una funzionalista, che distribuiscono bottigliette di aranciata per mille lire. La prima ha un unico stato interno non modificabile, S_0 , e fornisce l'aranciata se e solo se riceve 1000 lire. La seconda, a differenza della prima, possiede un'*organizzazione funzionale*: ha due stati interni, S_1 e S_2 , e fornisce l'aranciata se riceve una banconota da 1000 lire oppure due monete da 500 lire. Le due tavole di transizione sono le seguenti [cfr. Fodor 1981, 25]:

<i>Macchina comportamentista</i>		<i>Macchina funzionalista</i>		
	stato interno S_0		stato interno S_1	stato interno S_2
1000 lire intere	distribuisce aranciata	moneta da 500 lire	niente aranciata e passa in S_2	distribuisce aranciata e passa in S_1
		1000 lire intere	distribuisce aranciata e resta in S_1	distribuisce aranciata, 500 lire e passa in S_1

L'unico stato della macchina comportamentista, o - come la chiamano i cibernetici - "macchina banale", corrisponde alla sua "disposizione comportamentale"; ora, per quanto numerose possano essere le disposizioni comportamentali di un sistema, macchina o essere vivente, esse sono definite una volta per tutte e, dopo un sufficiente periodo di osservazione, diventano prevedibili. La macchina funzionalista, invece, ha stati interni, ovvero "stati funzionali", *interdefiniti*: il suo comportamento, a parità di "stimolo", dipende, in parte, dallo stato interno e lo stato interno dipende, in parte, dallo "stimolo". Così, per esempio, introducendo 500 lire si può avere o non avere l'aranciata, a seconda che la macchina sia in S_2 o in S_1 , ma essere in S_2 o in S_1 dipende dal denaro introdotto *precedentemente*⁶.

⁶Per questa ragione le macchine funzionaliste, chiamate dai cibernetici "macchine non-banali", dipendono dalla loro *storia* e sono perciò, considerate ad un certo istante, *imprevedibili* [von Foerster 1985, 131].

La seconda caratteristica importante delle due macchine, e *a fortiori* di quella funzionalista, è il fatto che esse potrebbero essere realizzate non solo in una molteplicità di materiali differenti, ma anche mediante innumerevoli dispositivi: «*la descrizione del software di una macchina distributrice di bibite non richiede logicamente che la sua realizzazione concreta avvenga mediante ingranaggi, leve e diodi*» [Fodor 1981, 24].

Fuori dalla metafora della macchina per le aranciate: i funzionalisti criticano l'identità mente-cervello perché, identificando la mente con il cervello, nega la mente a qualunque sistema tranne il cervello - in un certo senso, è troppo "sciovinista"; ma, d'altra parte, criticano il comportamentismo che, nell'attribuire la mente ad un sistema, non tiene conto dei suoi stati interni - è, in altre parole, troppo "liberale".

L'idea fondamentale del funzionalismo è questa: gli esseri umani possono essere concepiti come *macchine funzionaliste* i cui stati funzionali, i fenomeni mentali, sono *interdefiniti* e, almeno in linea di principio, *realizzabili* in una molteplicità di modi. In altre parole, *avere stati mentali* - per un sistema fatto di *qualunque* materiale - *significa possedere un tipo "appropriato" di organizzazione funzionale*. E ciò vale anche per noi esseri umani: *fossimo «anche fatti di emmenthal svizzero, la cosa non avrebbe nessuna importanza»* [Putnam 1975a, 318].

Se la natura della mente è funzionale, allora ciò che conta, per avere una mente, è solo il *software*, e non l'*hardware*. Da qui la famosa analogia: *la mente sta al cervello come i programmi stanno ai computer sui quali girano*. Ma l'analogia mente-programma è ancora più profonda. Un calcolatore non è altro che un sistema fisico di manipolazione simbolica: ad ogni stato dell'elaborazione corrispondono operazioni su dati codificati interpretabili semanticamente. Se si assume, come fanno i funzionalisti, che esistano cose come i *simboli mentali - rappresentazioni mentali con proprietà semantiche*, allora la metafora della mente come *software* diventa qualcosa di più di una semplice metafora: diventa un nuovo modo di considerare l'attività mentale. Come dice Fodor, pensare

un sistema (come il sistema nervoso) come un calcolatore vuol dire sollevare domande sulla natura del codice in cui quel sistema formula i suoi calcoli e sulle proprietà semantiche dei simboli nel codice. In effetti, l'analogia fra mente e calcolatore comporta proprio che si postulino simboli mentali [Fodor 1981, 27].

Questa è, in sostanza, la riflessione iniziale della cosiddetta “teoria computazionale della mente”. Per ora, essa ci interessa particolarmente perché introduce una nuova questione all’interno del problema mente-corpo. Come sappiamo, tale problema è costituito da una duplice domanda: qual è la natura della mente e quale il suo rapporto con il cervello? La teoria computazionale della mente suggerisce di ampliare il problema con la domanda: *e qual è il rapporto tra mente e computer?* E’ evidente che a seconda di cosa rispondiamo a quest’ultima domanda, dobbiamo ammettere o meno la possibilità, almeno teorica, che una macchina possa pensare.

Chiamata in causa dal problema *ampliato* mente-corpo, la comunità dei ricercatori che si occupano dell’IA ha risposto in modo non univoco. E ogni risposta ha ottenuto, a sua volta, consensi e dissensi da parte dei filosofi della mente. Alcune tra queste posizioni verranno esaminate nella parte IV della tesi; qui, invece, ci limitiamo a riportare una critica notevole che è stata avanzata contro la soluzione funzionalista al problema della natura della mente (lasciando da parte il problema del rapporto tra mente e computer).

Il problema, segnalato anche da Fodor, è quello dei “contenuti qualitativi” (*qualia*). Per mettere in luce questo problema, i filosofi si sono cimentati in una serie di esperimenti mentali cercando di immaginare situazioni in cui, *a parità di funzionamento, cervelli diversi fanno esperienze mentali diverse*. Se la natura della mente è funzionale, come sostengono i funzionalisti, ciò non dovrebbe essere possibile. Per comprendere gli esperimenti mentali che sono stati proposti, è necessario chiarire in qualche modo che cosa intendiamo per “contenuto qualitativo” di un’esperienza mentale. Fodor ci fornisce la seguente descrizione:

Immaginatevi di osservare una parete bianca attraverso un filtro rosso. Ora sostituite il filtro con un filtro verde e lasciate tutto il resto esattamente com’era. Quando cambia il filtro, cambia qualcosa nel carattere della vostra esperienza, e questo è il tipo di cosa che i filosofi chiamano contenuto qualitativo [Fodor 1981, 26].

Ciò premesso, consideriamo l’esperimento mentale dello “spettro invertito”. Dobbiamo immaginare due soggetti identici, anche psicologicamente, tranne che per il fatto che esperienze con il contenuto qualitativo “rosso” per uno dei due soggetti, hanno il contenuto qualitativo “verde” per l’altro. Ovviamente nel loro comportamento

nulla potrebbe rivelare la differenza, poiché *ambidue*, durante la loro infanzia, hanno imparato a chiamare «rosso» il colore dei pomodori maturi e delle albe fiammeggianti. Inoltre, è possibile che i due soggetti associno i loro *qualia* ad identici stati mentali: forse «*ambidue pensano la stessa cosa quando vedono i pomodori maturi, e si sentono depressi quando vedono il colore verde, e così via*» [Fodor 1981, 26]. Quali conclusioni si possono trarre da questo esperimento mentale? Conclusioni scettiche circa la natura funzionale della mente:

Sembrerebbe che i due osservatori possano condividere tutto ciò che si potrebbe confezionare entro la nozione del ruolo causale delle loro esperienze, e tuttavia il contenuto qualitativo delle esperienze possa essere tanto diverso quanto si voglia. Se questo è possibile, allora il resoconto del funzionalista non si adatta a stati mentali che abbiano un contenuto qualitativo. Se una persona ha una esperienza di verde mentre l'altra ha una esperienza di rosso, allora sicuramente le due persone debbono essere in stati mentali diversi [*ibid.*].

Un altro esperimento mentale molto famoso, anch'esso relativo al problema dei *qualia*, è stato proposto da Ned Block nel 1980. Se la mente è il *software* del cervello, e se un programma può, almeno in teoria, essere descritto funzionalmente da una tavola di transizione del tipo di quella della macchina distributrice di bibite, allora possiamo immaginare la seguente situazione: così come due persone, S_1 e S_2 , possono realizzare l'organizzazione funzionale di una macchina distributrice di bibite⁷, organizzando in modo "appropriato" la popolazione della Cina si potrebbe far sì che essa realizzi la specificazione funzionale di uno stato mentale (del quale, nell'esperimento mentale, conosciamo l'immensa organizzazione funzionale). Ciò però non sembra molto convincente: è difficile credere che la popolazione della Cina, per quanto organizzata, possa mai divenire *essa stessa* un soggetto di esperienza e provare davvero delle sensazioni [Bechtel 1988, 215].

Di fronte all'obiezione dei *qualia* sono state proposte molte versioni modificate *ad hoc* del funzionalismo; per ovvie ragioni di spazio, non possiamo considerarle, né possiamo dilungarci ulteriormente sul funzionalismo. Più avanti, tuttavia,

⁷ S_1 e S_2 ricevono a turno il denaro. Se è il turno di S_1 e l'utente introduce nel sistema una moneta da 500 lire, S_1 mette in cassa la moneta e lascia il posto a S_2 ; se è il turno di S_1 e l'utente introduce 1000 lire intere, allora S_1 prende dal contenitore delle aranciate una bibita, la fornisce all'utente e rimane al suo posto; se è il turno di S_2 e l'utente introduce una moneta da 500 lire, S_2 distribuisce un'aranciata e lascia il posto a S_1 ; se infine è il turno di S_2 e l'utente introduce 1000 lire intere, allora S_2 prende un'aranciata dal contenitore e una moneta da 500 lire dalla cassa, fornisce entrambe all'utente e lascia il posto a S_1 .

riprenderemo alcuni dei temi toccati in questo paragrafo e approfondiremo la teoria computazionale della mente, sostenuta tanto dai funzionalisti quanto dai ricercatori dell'IA convenzionale.