Biblioteca Essenziale

CL 20-7065-0

Il libro

La mente. La sua architettura. Le sue funzioni e disfunzioni. Un nuovo modo di affrontare questi temi, intrecciando filosofia e psicologia.



L'autore

Massimo Marraffa svolge attività di ricerca presso il Dipartimento di Filosofia dell'Università Roma Tre.

ISBN 88-420-7065-3

Biblioteca Essenziale

Filosofia della psicologia



Massimo Marraffa

GLF Editori Laterza

Filosofia

serie diretta da
Tito Magri

VOLUMI PUBBLICATI

Maurizio Ferraris ~ L'ermeneutica
Salvatore Veca ~ La filosofia politica
Eva Picardi ~ Le teorie del significato
Michele Di Francesco ~ La coscienza
Piergiorgio Donatelli ~ La filosofia morale
Diego Marconi ~ Filosofia e scienza cognitiva
Nicla Vassallo ~ Teoria della conoscenza

Massimo Marraffa

Filosofia della psicologia

GLF *Editori Laterza*

© 2003, Gius. Laterza & Figli

Prima edizione 2003

Proprietà letteraria riservata
Gius. Laterza & Figli Spa,
Roma-Bari

Finito di stampare
nel luglio 2003
Poligrafico Dehoniano -
Stabilimento di Bari
per conto della
Gius. Laterza & Figli Spa
CL 20-7065-0
ISBN 88-420-7065-3

È vietata la riproduzione, anche parziale, con qualsiasi mezzo effettuata, compresa la fotocopia, anche ad uso interno o didattico. Per la legge italiana la fotocopia è lecita solo per uso personale *purché non danneggi l'autore*. Quindi ogni fotocopia che eviti l'acquisto di un libro è illecita e minaccia la sopravvivenza di un modo di trasmettere la conoscenza. Chi fotocopia un libro, chi mette a disposizione i mezzi per fotocopiare, chi comunque favorisce questa pratica commette un furto e opera ai danni della cultura.

a Tiziana

Premessa

Nel corso degli ultimi anni la filosofia della scienza è diventata sempre più 'locale', spostando l'attenzione dalle caratteristiche generali della prassi scientifica agli apparati concettuali delle singole discipline scientifiche. La *filosofia della psicologia* è uno dei risultati di questa crescente specializzazione.

La psicologia di cui il filosofo della scienza si occupa è la *psicologia cognitiva* o *cognitivismo*, la disciplina che studia i processi di elaborazione di informazioni negli organismi complessi. Ciò che rende questa scienza peculiare è il suo essere sospesa tra due mondi. Da un lato, l'immagine ordinaria di noi stessi in quanto *persone*, vale a dire in quanto soggetti di esperienze coscienti, stati intenzionali e agire deliberato (la cosiddetta *psicologia del senso comune*). Dall'altro, la sfera *subpersonale* degli eventi cerebrali, oggetto della neuroscienza. Pertanto uno dei compiti principali del filosofo della psicologia è venire a capo di questa peculiarità, tentando di fare un po' di luce sulle relazioni che legano questi differenti modi in cui descriviamo noi stessi.

La ricerca che ha condotto a questo libro si è svolta nell'ambito del programma di ricerca Cofin 2001-2003 sul tema *Causalità e intenzionalità* (coordinatrice Rosaria Egidi).

Nel corso di questa ricerca ho potuto trarre profitto da discussioni con Cristiano Castelfranchi, Michele Di Francesco, Rosaria Egidi, Simone Gozzano, Diego Marconi, Giacomo Marramao, Cri-

stina Meini, Carmela Morabito, Sandro Nannini, Domenico Parisi e Roberto Pujia.

Mario De Caro, Marica De Vincenzi, Francesco Ferretti, Pierdaniele Giaretta, Giovanni Jervis, Tito Magri, Alfredo Paternoster, Laura Sparaci e Luca Surian hanno avuto la pazienza di leggere e commentare il testo, aiutandomi a correggere errori e a migliorare l'esposizione.

Filosofia della psicologia

1

Il progetto computazionale in psicologia

Si sente spesso ripetere che la nascita della psicologia scientifica ha coinciso con il suo distacco dalla filosofia. In realtà un simile distacco non è mai avvenuto. Di certo non ha avuto luogo ad opera della psicologia cognitiva, un programma di ricerca fondato su una sofisticata variante delle teorie rappresentazionali della mente che nel corso dei secoli si sono susseguite in seno alla tradizione empirista. Il comune denominatore di queste teorie è l'idea che la relazione tra mente e mondo è mediata da oggetti interni dotati di proprietà semantiche e causali: 'idee' nella terminologia filosofica in voga tra il XVII e il XIX secolo; 'concetti' e 'rappresentazioni mentali' nel vocabolario dei cognitivisti contemporanei. Dove la psicologia cognitiva si è allontanata dalla tradizione empirista è stato nell'attribuire a questi oggetti interni una struttura sintattica, e dunque nell'ipotizzare che le loro relazioni causali abbiano una natura *computazionale* piuttosto che *associativa*. Ciò ha consentito per la prima volta di sviluppare con rigore la celebre tesi hobbesiana del *pensiero come calcolo*.

Questo progetto computazionale in psicologia e il rapporto che esso intrattiene con il mentalismo del senso comune sono i temi del presente capitolo.

Seduto su una sedia, davanti al computer, le mie dita percorrono sicure la tastiera. A un tratto smetto di scrivere, mi alzo per dirigermi verso la cucina, dove apro il frigorifero per estrarne una lattina di birra. Se a qualcuno fosse capitato di osservare questa scena, di certo non l'avrebbe trovata insolita. Sicuramente io, ripensando a questo episodio, non ne sono minimamente sorpreso. Infatti, una semplicissima spiegazione è a portata di mano. Mentre ero intento a scrivere, ho avvertito un *bisogno*, avevo sete, da cui ha preso corpo il *desiderio* di una birra. A quel punto, in base alla *credenza* che una lattina di birra fosse conservata nel frigorifero, ho pianificato l'azione che mi ha consentito di estinguere il bisogno.

Una sequenza di azione è stata dunque spiegata in base all'attribuzione di alcuni stati mentali. Questa spiegazione può essere costruita sia alla *prima persona* (attribuisco gli stati mentali a me stesso) sia alla *terza persona* (l'osservatore mi attribuisce gli stati mentali). Inoltre l'auto- oppure l'etero-attribuzione di stati mentali avrebbe potuto costituire la base anche di una *previsione* della sequenza di azione (sapendo che desidero una birra e che credo che ve ne sia una in frigorifero, è possibile inferire che mi alzerò per recarmi in cucina).

1.1. *La teoria della teoria*

Ognuno di noi, nell'interazione quotidiana con se stesso e con gli altri, è impegnato senza sosta nella costruzione di spiegazioni e di previsioni fondate sull'attribuzione di stati mentali. Secondo quella che Morton (1980) ha definito *teoria della teoria*, questa pratica mentalistica poggia su una teoria o, per meglio dire, una proto-teoria spontanea: la cosiddetta *psicologia del senso comune* o *popolare* (*folk psychology*). Quest'ultima avrebbe natura teorica nel senso di essere costituita da un complesso di conoscenze integrate e coerenti che mettono ordine nella proteiforme sfera del mentale in base a due categorie: gli stati soggettivi dell'esperienza cosciente – spesso denominati *qualia* – e gli stati *intenzionali*.

I *qualia* sono le proprietà esperienziali o introspettive degli stati mentali. La loro essenza sembra consistere nell'essere colte da un punto di vista soggettivo o alla prima persona: percepire una certa tonalità di rosso *fa un certo effetto*; quando si crede di aver subito un torto *si prova qualcosa*. Nel loro insieme, queste entità mentali definiscono l'ambito della *coscienza fenomenica*[1].

Gli stati intenzionali (credenze, desideri, intenzioni, speranze ecc.) sono stati che godono della proprietà di *vertere su*, o di *essere diretti verso*, qualcosa. Ad esempio, se io credo che ci sia una birra nel frigorifero, la mia credenza è diretta verso un *oggetto* o *contenuto intenzionale*, vale a dire ciò che è espresso dall'enunciato 'c'è una birra in frigorifero'. Spesso gli stati intenzionali sono definiti *atteggiamenti proposizionali* in quanto – come facilmente si rileva dall'esempio appena fatto – nell'attribuirli a un soggetto ci avvaliamo di enunciati della forma '*S* crede (o desidera/intende/spera ecc.) che *p*', dove la proposizione *p* esprime il contenuto dello stato mentale del soggetto *S*.

Forte di questo vocabolario mentalistico, lo psicologo del senso comune formula conoscenze organizzabili in un insieme di *generalizzazioni nomologiche* (ossia 'che hanno carattere di leggi'). Consideriamo un'altra elementare sequenza di azione: Mario è uscito di casa portando con sé l'ombrello. Una tipica spiegazione in termini di atteggiamenti proposizionali potrebbe essere: Mario è uscito portando con sé l'ombrello perché credeva che stesse piovendo, desiderava non bagnarsi ed era convinto di poter soddisfare tale desiderio utilizzando un ombrello. L'espressione 'perché' indica che la *causa* del comportamento di Mario viene ricondotta ai suoi atteggiamenti di credenza e desiderio. Si tratta di un tipo di

[1] Block (1995) distingue questo significato del termine 'coscienza' da quello utilizzato in psicologia scientifica: la coscienza *di accesso* (o *cognitiva*, come la chiama Chalmers 1996). Mentre la coscienza fenomenica rinvia a fenomeni in prima persona caratterizzati da una qualità intrinseca, la coscienza di accesso è un costrutto teorico introdotto dallo psicologo per dar conto della capacità di un sistema di aver accesso ai propri stati interni ai fini (fra l'altro) del ragionamento, della pianificazione dell'azione e dell'esplicitazione verbale.

relazione causale che viene catturata da una generalizzazione nota come *sillogismo pratico*:

> se *X* desidera che *p*, e *X* crede che *p* se *q*,
> allora *X* farà in modo di causare *q*
> (se Mario desidera non bagnarsi, e crede che portarsi dietro l'ombrello servirà allo scopo, allora uscirà con l'ombrello).

Questa generalizzazione specifica una relazione causale tra certi stati mentali e un comportamento manifesto. Altre generalizzazioni psicologico-popolari sussumono interazioni causali fra stimoli sensoriali e stati mentali: ad esempio, 'se un oggetto rosso è posto in piena luce di fronte a un soggetto, questi avrà una percezione di rosso'. Altre ancora specificano le relazioni causali tra stati mentali e altri stati mentali: ad esempio, 'se un individuo crede che *p*, e crede che se *p* allora *q*, allora crederà *q*'[2].

1.2. *Lo statuto epistemico della psicologia del senso comune*

Gli psicologi sociali hanno indagato la nostra capacità di attribuire stati mentali almeno a partire dagli anni Quaranta. In un classico esperimento di Heider e Simmel (1944) veniva mostrato ad alcuni soggetti un breve filmato in cui tre figure geometriche si muovevano in uno spazio delimitato, attorno a un rettangolo avente una parte di lato mobile (Fig. 1.1).

Interrogati su quanto avevano visto, quasi tutti i soggetti interpretarono i movimenti delle tre forme in chiave antropomorfica – ad esempio, quando il lato mobile del rettangolo si spostava in seguito al contatto con le figure, l'evento veniva descritto come se si trattasse di un'azione umana, vale a dire l'apertura di una porta. Questi risultati suggerivano la presenza di una capacità di attribuzione di stati mentali universale e in larga misura automatica.

Lo sviluppo di tali ricerche condurrà al libro *The Psychology of Interpersonal Relations* di Heider (1958), un'opera che costituisce il principale referente storico nello studio della psicologia del senso comune. In particolare, da questo lavoro trae origine la *teoria dell'attribuzione causale*, il settore della psicologia sociale che indaga i meccanismi che presiedono alla costruzione delle spiegazioni ordinarie sulle cause del nostro e altrui comportamento.

[2] Queste generalizzazioni sono sempre limitate da clausole *ceteris paribus*, su cui vedi *infra*, § 1 del cap. 2.

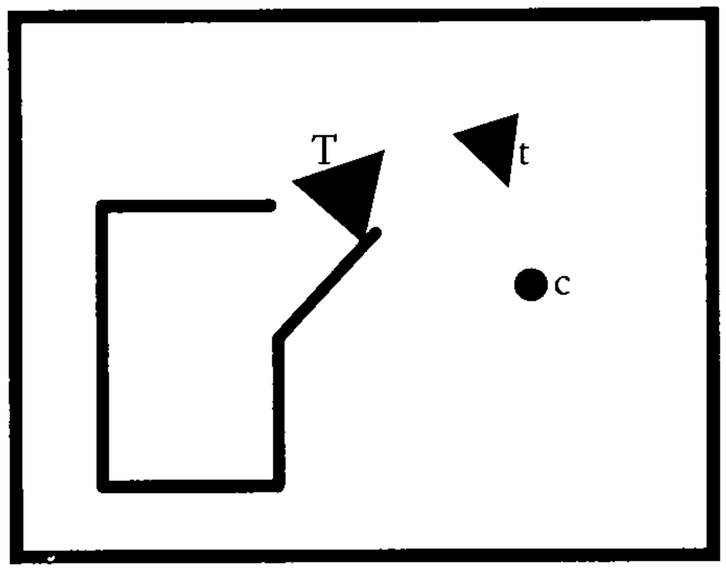


Figura 1.1. Un cerchio (c) e due triangoli, uno più grande (T) e l'altro più piccolo (t), si muovono attorno a un rettangolo con un lato mobile (da Heider e Simmel 1944).

Il lavoro di Heider e la teoria dell'attribuzione esemplificano chiaramente il contrasto tra due differenti modi di accostarsi alla psicologia del senso comune. Il primo prese tale psicologia prevalentemente *in positivo*, giudicando i concetti del senso comune un buon punto di partenza per sviluppare una teoria scientifica del comportamento sociale. Invece la maggior parte delle ricerche sull'attribuzione causale è stata fedele alla lezione metodologica del comportamentismo, prendendo la psicologia del senso comune prevalentemente *in negativo* e compiendo essenzialmente un'analisi dell'accuratezza e dei limiti delle spiegazioni ordinarie del comportamento sociale (Jervis 1993, pp. 52-54).

Nella filosofia della psicologia degli ultimi trent'anni il

contrasto tra questi due opposti atteggiamenti nei confronti dello statuto epistemico della psicologia del senso comune ha plasmato il dibattito sul destino dell'immagine ordinaria di noi stessi alla luce dell'imponente sviluppo delle scienze della mente. Qui l'interrogativo fondamentale è: le entità teoriche postulate dalla psicologia del senso comune troveranno posto nell'ontologia di una scienza matura della mente? E le risposte vanno da quella affermativa di Jerry Fodor che prende 'in positivo' la teoria psicologica popolare, giudicandola l'unica base possibile su cui edificare una psicologia scientifica; a quella pessimistica di Stephen Stich, che sostanzia i suoi dubbi nei confronti del concetto popolare di credenza attingendo, fra l'altro, proprio alle ricerche sull'attribuzione causale; a quella ancor più pessimistica di Patricia e Paul Churchland che considerano l'emancipazione della psicologia dai concetti del senso comune la condizione della sua scientificità, intesa da questi studiosi come 'riducibilità' alla neuroscienza.

L'analisi di queste due prospettive concernenti la validità della teoria psicologica popolare – la prima 'conciliabilista', l'altra 'eliminativista' – ci impegnerà nel corso di tutto il libro. Nei parr. 2 e 3 presenteremo la nascita della psicologia cognitiva come la risultante del ripudio delle istanze eliminativiste del comportamentismo in favore di un progetto conciliabilista che costituisce una sorta di 'mentalismo sperimentale'. Nel par. 4 vedremo però che dubbi eliminativisti non hanno mai smesso di assillare i cognitivisti (specialmente in psicologia sociale e in neuropsicologia cognitiva). Tanto che da queste inquietudini ha infine preso corpo un nuovo eliminativismo fondato su un importante sviluppo della psicologia cognitiva: il *connessionismo*. Questo è il tema di cui ci occuperemo nel cap. 4.

2. *Dentro la scatola nera: le vicissitudini dell'informazione*

Gran parte della psicologia filosofica classica, di orientamento sia razionalista che empirista, ha sottoscritto l'idea di un accesso privilegiato del soggetto ai contenuti e alle operazioni

della mente. Descartes considerava la mente come pienamente trasparente a se stessa e accessibile in tutte le sue proprietà significative a una facoltà di introspezione o riflessione cosciente. Una facoltà che in seguito i filosofi empiristi avrebbero caratterizzato come una forma di percezione interna.

Alla luce di questo ottimismo nei riguardi della conoscenza alla prima persona, non sorprende che tra la fine del XIX secolo e i primi decenni del XX la psicologia scientifica si proponga come una psicologia della coscienza fenomenica. Confidando nella possibilità di rendere l'introspezione un metodo di indagine rigoroso, che avrebbe elevato la psicologia al rango delle altre scienze della natura, i primi psicologi sperimentali esplorano la fenomenologia della mente umana, scandagliano minuziosamente lo spazio dei contenuti della coscienza, si sforzano di offrire una descrizione esaustiva del paesaggio mentale come appare al soggetto.

2.1. *Il comportamentismo eliminativista*

Dato il comune idioma mentalistico, questi psicologi introspezionisti non si sarebbero trovati in difficoltà a parlare con poeti, critici letterari, storici ed economisti: «Un lettore non specialista del 1910 si sarebbe trovato in un territorio familiare sia studiando i *Principles of Psychology* di William James sia leggendo i romanzi del fratello Henry» (Stich 1983, p. 1). Il comportamentismo – almeno nella sua forma più radicale – pose fine ai buoni rapporti tra psicologia scientifica e psicologia del senso comune. Thorndike, Pavlov, Watson si proposero di abbandonare il vocabolario mentalistico, bollandolo come un incomprensibile gergo prescientifico, da collocare sullo stesso piano della «superstizione», della «magia» e delle pratiche «vudù» (Watson 1930, pp. 2-5). Di conseguenza una psicologia che aspirasse a essere scientifica doveva far conto esclusivamente su dati osservabili, e questi erano gli *stimoli* (eventi fisici nell'ambiente) a cui un soggetto è sottoposto e le sue *risposte* (comportamento pubblicamente osservabile) a tali stimoli. Ne derivò una concezione della spie-

gazione psicologica davvero austera: lo psicologo, munito soltanto della teoria del condizionamento di Pavlov e della legge dell'effetto di Thorndike, doveva limitarsi a registrare le connessioni associative tra classi di stimoli e classi di risposte. Quel che accadeva dentro la 'testa' fra l'input e l'output era oggetto dell'indagine fisiologica: l'organismo era concepito come una *scatola nera*.

Nella misura in cui espunge la mente dalla spiegazione psicologica e dall'ontologia della scienza del comportamento, il comportamentismo è una variante della dottrina nota come *eliminativismo*: la previsione che i concetti mentalistici del senso comune scompariranno dalla psicologia scientifica in modo simile a quanto è avvenuto in passato, quando la ricerca scientifica ha portato all'abbandono della teoria popolare della stregoneria o della teoria protoscientifica del calorico. Più precisamente, l'eliminativista costruisce un'argomentazione che dalla premessa asserente la falsità della teoria psicologica popolare conduce alla conclusione che le entità da essa postulate non esistono, nello stesso modo in cui non esistono le streghe e il calorico. La versione comportamentista dell'eliminativismo prevede che la teoria scientifica che prenderà il posto della teoria psicologica del senso comune sarà espressa nel vocabolario del comportamento fisico.

Il comportamentismo eliminativista è un tema ricorrente negli scritti di Watson (1930) e di Skinner (1984), anche se talvolta questi studiosi oscillano tra un'interpretazione eliminativista del comportamentismo (una tesi ontologica ed esplicativa: gli stati mentali non esistono e dunque la spiegazione del comportamento animale non avrà carattere mentalistico) e altre due interpretazioni: la tesi *metodologica* che gli stati mentali esistono, ma sono irrilevanti per la prassi scientifica; e la tesi *semantica*, legata ai nomi di Rudolf Carnap e Carl G. Hempel, secondo cui tutti gli asserti contenenti termini psicologici sono *traducibili* in asserti contenenti esclusivamente termini che si riferiscono al comportamento fisico. Quest'ultima tesi – nota come comportamentismo *analitico* o *logico* – è un progetto *riduzionistico*: ossia gli stati mentali non sono eliminati, bensì *identificati con* disposizioni comportamentali.

2.2. *Mappe cognitive e strutture della sintassi*

Negli anni Trenta e Quaranta la crescente consapevolezza dei limiti dello schema stimolo-risposta fa evolvere questo comportamentismo fortemente oggettivistico, interessato esclusivamente al comportamento pubblicamente osservabile, verso quello che sarà poi, a partire dagli anni Sessanta, il cognitivismo. Una pietra miliare in questa evoluzione è il lavoro sulle *mappe cognitive* del neocomportamentista Edward Tolman. In un celebre esperimento lo studioso pose un ratto nel labirinto illustrato in Fig. 1.2.

Il ratto iniziava a percorrere il labirinto dal punto *P*. Dopo una prima fase in cui riceveva il cibo su tutte e tre le piattaforme, il ratto veniva ripetutamente rinforzato nel punto *A*, situato alla destra di *P*. In tal modo il ratto imparava a percorrere il tratto da *P* ad *A*. Di che tipo di apprendimento si trattava? Di un apprendimento puramente motorio, come sarebbe stato legittimo attendersi alla luce del paradigma del condizionamento operante? Oppure di un apprendimento della posizione del cibo rispetto all'intero piano del percorso? In altre parole, il ratto aveva appreso un semplice schema

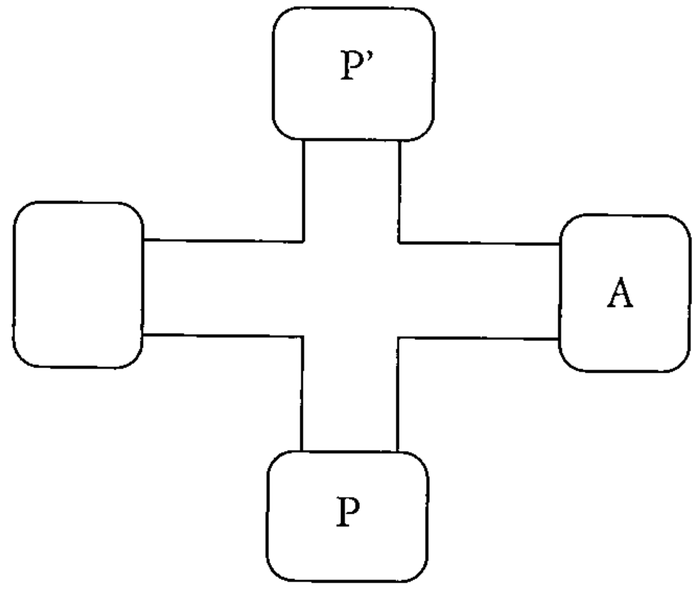


Figura 1.2. Labirinto radiale (da Mazzoni 2000).

motorio (gira a destra) oppure aveva appreso le relazioni spaziali tra i vari elementi costituenti il percorso, e i suoi movimenti dipendevano da *una mappa del labirinto rappresentata internamente*? Per dirimere la questione, Tolman collocò il ratto in una nuova posizione di partenza *P'*, rispetto alla quale il punto *A* in cui riceveva il rinforzo si trovava non più a destra ma a sinistra. Due i possibili scenari. Se il ratto si era limitato ad apprendere la risposta 'gira a destra', anche in questa nuova situazione si sarebbe dovuto dirigere verso destra. Se invece aveva appreso una mappa del labirinto, nella nuova situazione avrebbe dovuto muoversi verso sinistra, dove era situato il cibo. L'esperimento confermò il secondo scenario: il ratto si dirigeva direttamente verso il cibo, confermando così l'ipotesi dell'esistenza di una rappresentazione interna del labirinto. Di conseguenza la capacità di 'navigazione spaziale' del ratto non era spiegabile in base all'instaurarsi di un'associazione tra certi stimoli e certe specifiche risposte muscolari, richiedendo bensì l'ipotesi che gli animali si costruissero «all'interno del sistema nervoso, adattamenti funzionali che operano come mappe cognitive» (Tolman 1948, p. 193).

Vi furono alcuni ingegnosi tentativi di potenziare lo schema stimolo-risposta per dar conto di questi e altri risultati sperimentali di Tolman senza le sue concessioni mentalistiche (vedi ad esempio Hull 1943). Ma tale schema si rivelò completamente inefficace quando dal comportamento dei ratti in un labirinto si passò a considerare i complessi comportamenti caratteristici della nostra specie. L'uso del linguaggio è un buon esempio. E non è dunque un caso che uno dei principali fattori del trapasso dal comportamentismo al cognitivismo sia stato l'impetuoso sviluppo, a partire dalla fine degli anni Cinquanta, di una teoria mentalistica del linguaggio: la *linguistica generativa* di Noam Chomsky.

Fra i molti argomenti volti a mostrare la necessità di spiegare l'uso del linguaggio in termini di strutture e processi interni, Chomsky ne formulò uno destinato a diventare un 'ferro del mestiere' dei cognitivisti: l'argomento della *povertà dello stimolo*. L'argomento ci invita a prendere in esame l'input

e l'output del processo di acquisizione del linguaggio. Se l'output contiene in misura significativa informazioni che non sono riscontrabili nell'input, l'unica ipotesi plausibile è che il sovrappiù di informazione sia un contributo del parlante/ascoltatore di una lingua naturale. Chomsky presentò prove empiriche cogenti a riprova dell'enorme scarto tra lo stato finale dell'acquisizione del linguaggio (la conoscenza di una struttura grammaticale estremamente complessa) e gli input 'inquinati e corrotti' ricevuti dall'esterno (le frasi prodotte dai membri di una determinata comunità linguistica). Il sovrappiù di informazione contribuito dal parlante/ascoltatore doveva essere una base innata di regole (la *grammatica universale*) in grado di determinare *a priori* le strutture grammaticali possibili (su cui vedi *infra*, § 2.1 del cap. 2).

2.3. *L'uomo elaboratore di informazioni*

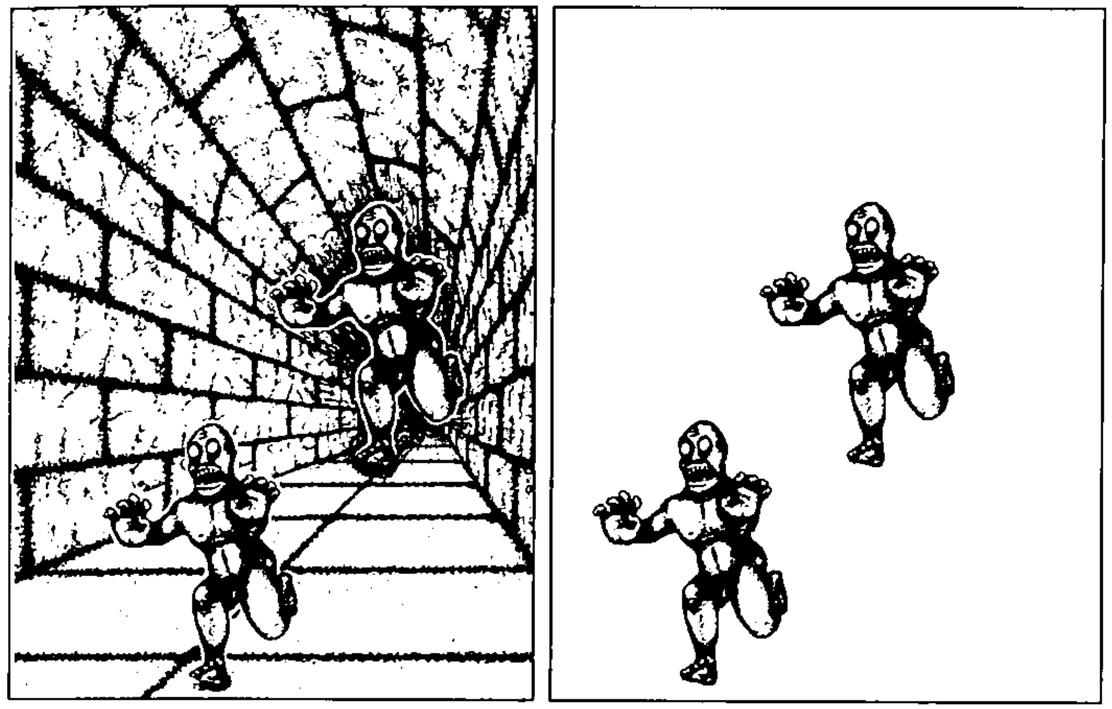
«Il cognitivismo moderno prende avvio dagli argomenti della povertà dello stimolo» ha scritto Fodor (2001, p. 61). Se l'output è più ricco dell'input, deve essere intervenuta un'elaborazione interna di qualche tipo, una trasformazione che ha modificato lo stimolo. Questo lavoro di trasformazione compiuto dall'organismo è una causa *inosservabile* che il cognitivista inferisce dal comportamento.

Siamo finalmente in condizione di definire la psicologia cognitiva. Essa è la scienza che studia i processi di elaborazione di informazioni negli organismi complessi, ossia «tutti quei processi per mezzo dei quali l'input sensoriale viene trasformato, ridotto, integrato, immagazzinato, recuperato e infine utilizzato» (Neisser 1967, p. 1). Scaviamo un po' in questa celebre definizione. L'organismo di cui si occupa il cognitivista non è più 'vuoto' come lo voleva il comportamentismo. Tra la stimolazione degli organi di senso e l'emissione della risposta vi è nuovamente la *mente*, concepita come 'elaboratore di informazioni'. Le informazioni in ingresso vengono *codificate* nella mente, divenendo in tal modo *oggetti interni* (le *rappresentazioni mentali*) suscettibili di elaborazioni di vario

tipo. In particolare, questi oggetti possono essere *trasformati*: la nostra rappresentazione della realtà è il frutto non già di una registrazione passiva dell'ambiente fisico, ma di una costruzione attiva che può dar luogo sia a *riduzione* che a *integrazione*. La nostra capacità di prestare attenzione è limitata a un numero relativamente modesto di stimoli fisici; e una quantità ancora più modesta può venire ricordata. È dunque possibile che parte dell'informazione in ingresso vada perduta, e in questo caso ha luogo una riduzione. Ma soprattutto l'input sensoriale può essere integrato; ed è qui che entrano in gioco celebri argomenti della povertà dello stimolo.

Prendiamo, ad esempio, le *costanze percettive*. Lo stimolo *prossimale* (vale a dire l'informazione disponibile nell'immagine retinica) varia continuamente al variare della posizione reciproca tra l'osservatore e l'oggetto osservato (lo stimolo *distale*). Ad esempio, l'immagine retinica di un oggetto si rimpiccolisce man mano che l'osservatore si allontana, e si ingrandisce al suo avvicinarsi. Ciò malgrado noi non vediamo gli oggetti ingrandirsi e rimpiccolirsi, li vediamo sempre all'incirca della stessa grandezza ma a distanze diverse: un fenomeno noto come *costanza di grandezza*. In questo caso, come in quello del linguaggio, non è possibile ridurre l'output all'input. E dal momento che l'output è sottodeterminato dall'input, l'organismo percipiente deve aver apportato un contributo di informazioni, operato una qualche integrazione percettiva. Secondo una teoria formulata per la prima volta da Hermann von Helmholtz, la costanza di grandezza è il prodotto di un'*inferenza inconscia*: il sistema percettivo compie una sorta di ragionamento per cui quando vede un oggetto e 'sa' che questo è lontano, aggiusta la grandezza percepita ingrandendola; e quando 'sa' che è vicino, aggiusta la grandezza rimpiccolendola (Fig. 1.3).

Nel ricostruire le vicissitudini dell'informazione tra gli stimoli sensoriali e la risposta comportamentale, il cognitivista analizza i processi di trasformazione, immagazzinamento, recupero e uso di informazioni, *scomponendoli* in gruppi di operazioni distinte, detti *stadi di elaborazione*. Queste elaborazioni sono concepite come *calcoli* o, come d'ora in poi diremo, *computazioni*, vale a dire sequenze ordinate di operazioni condotte, in conformità a *regole*, su strutture dati (le rappresentazioni mentali) che mediano le risposte comportamentali di un organismo agli stimoli sensoriali.



(a) (b)

Figura 1.3. Il meccanismo di costanza di grandezza all'opera in un'illusione creata da Roger Shepard. La creatura sullo sfondo e quella in primo piano nella figura (a) hanno identiche dimensioni (come mostra la figura b), e quindi la medesima proiezione retinica. Tuttavia due oggetti posti a distanze diverse possono avere la stessa proiezione retinica solo se l'oggetto più distante è più grande. Il meccanismo della costanza di grandezza compensa l'effetto apparente della distanza, accrescendo le dimensioni della creatura sullo sfondo.

Nella misura in cui caratterizza la mente in termini computazionali, la psicologia cognitiva è parte integrante della *scienza cognitiva*, vale a dire quel progetto di studio interdisciplinare dell'intelligenza naturale e artificiale che inizia a maturare alla fine degli anni Cinquanta, finendo con l'assumere un assetto intellettuale e istituzionale stabile agli inizi degli anni Ottanta (la cosiddetta scienza cognitiva *classica*).

Questa maturazione non è stata indolore. La gestazione della scienza cognitiva si era caratterizzata per una coopera-

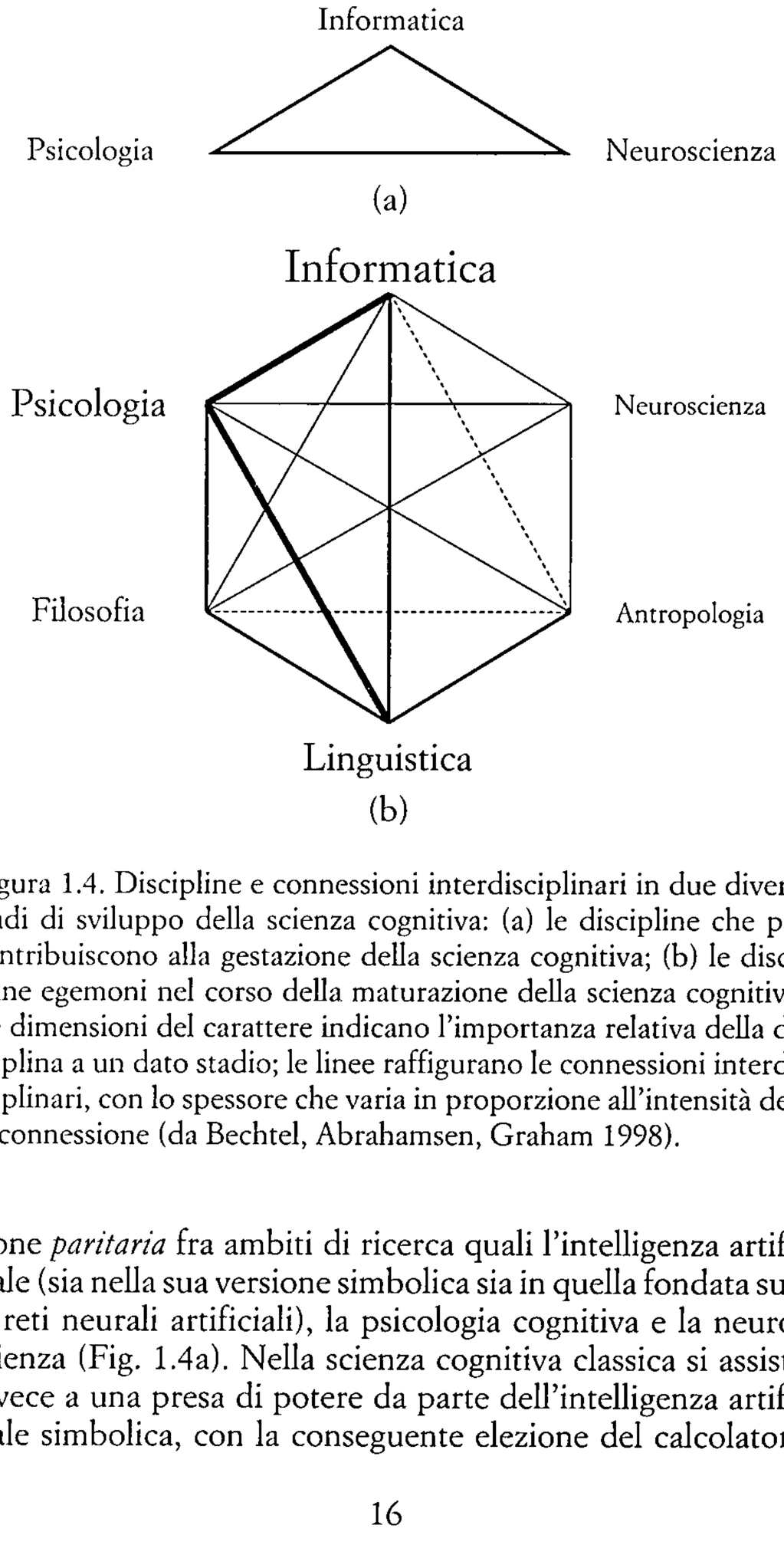


Figura 1.4. Discipline e connessioni interdisciplinari in due diversi stadi di sviluppo della scienza cognitiva: (a) le discipline che più contribuiscono alla gestazione della scienza cognitiva; (b) le discipline egemoni nel corso della maturazione della scienza cognitiva. Le dimensioni del carattere indicano l'importanza relativa della disciplina a un dato stadio; le linee raffigurano le connessioni interdisciplinari, con lo spessore che varia in proporzione all'intensità della connessione (da Bechtel, Abrahamsen, Graham 1998).

zione *paritaria* fra ambiti di ricerca quali l'intelligenza artificiale (sia nella sua versione simbolica sia in quella fondata sulle reti neurali artificiali), la psicologia cognitiva e la neuroscienza (Fig. 1.4a). Nella scienza cognitiva classica si assiste invece a una presa di potere da parte dell'intelligenza artificiale simbolica, con la conseguente elezione del calcolatore

digitale a fonte d'ispirazione primaria per lo studio della mente. La scienza cognitiva viene così intesa come lo studio dei processi mentali in quanto elaborazioni di informazioni da parte di un calcolatore, naturale o artificiale. In questa prospettiva, essa è un programma di ricerca che orienta pressoché esclusivamente ricerche psicologiche, linguistiche e informatiche: la psicologia cognitiva e la linguistica (intesa chomskianamente come parte della psicologia) conducono indagini su un particolare tipo di calcolatore, quello 'umano'; l'intelligenza artificiale, invece, ha come oggetto di studio il calcolatore 'artificiale' (Fig. 1.4b).

La filosofia a orientamento naturalistico ha svolto un ruolo importante nella genesi e nell'evoluzione del progetto computazionalista in psicologia. Essa infatti ne ha portato alla luce e gradualmente raffinato i fondamenti teorici, che sono essenzialmente due. L'idea che l'identità degli stati mentali è determinata dal ruolo causale che essi svolgono nella vita cognitiva di un agente, indipendentemente dal modo in cui tale ruolo è realizzato fisicamente (*funzionalismo*). E la tesi secondo cui i processi mentali sono computazioni, vale a dire manipolazioni di rappresentazioni simboliche governate da regole sintattiche (*teoria computazionale e rappresentazionale della mente*).

3. *Il funzionalismo computazionale e rappresentazionale*

La psicologia che sottoscrive il progetto della scienza cognitiva classica non è una riproposizione del mentalismo introspezionista. Come vedremo fra breve, la mente di cui si occupano molte ricerche di psicologia cognitiva è una sorta di formalizzazione di quella parte della psicologia del senso comune che riguarda gli atteggiamenti proposizionali (la psicologia *intenzionale*). La psicologia della coscienza fenomenica ne resta dunque esclusa. Ciò non sarebbe stato possibile nell'universo concettuale prefreudiano, in cui si riteneva che i problemi della coscienza e dell'intenzionalità fossero intrinsecamente collegati – che il pensiero fosse, in quanto tale, co-

sciente, e la coscienza fosse, in quanto tale, coscienza di qualche oggetto intenzionale. Freud ha mutato questo stato di cose nel momento in cui ha conferito plausibilità all'idea che la spiegazione del comportamento richiede di ipotizzare stati intenzionali *inconsci*. Infatti,

> nel corso di questo secolo, soprattutto nella linguistica chomskiana e in psicologia cognitiva, la tesi freudiana è stata ampiamente confermata. *Divide et impera* – concentrarsi sull'intenzionalità e ignorare la coscienza – si è rivelata finora una strategia di ricerca coronata da notevole successo (Fodor 1991, p. 12).

In realtà, qui si può essere più radicali di Fodor, sostenendo che la psicologia cognitiva non si è limitata a confermare Freud, ma è andata oltre. Infatti, come ha fatto osservare Jervis, il concetto freudiano di inconscio si definisce «'per differenza' (e per certi versi anche in subordine) rispetto a un concetto di coscienza preso da Freud come 'dato' ed evidente» (Jervis 1999, pp. 21-22). La psicologia cognitiva opera invece un «rovesciamento rafforzativo degli interrogativi psicodinamici tradizionali», prendendo le mosse dal chiedersi non già come sia possibile l'inconscio, ma come sia mai possibile la coscienza (Jervis 1993, p. 301). Così facendo, il cognitivista accoglie in pieno l'impostazione metodologica darwiniana che, assumendo la continuità tra la mente animale e quella umana, prescrive di portare avanti lo studio della coscienza ricostruendo il processo che dalle funzioni psicologiche più elementari (indagabili nell'animale e nell'infante) conduce a quelle funzioni psicologiche più complesse che rendono possibile la mente autocosciente adulta. In questa prospettiva, l'esperienza soggettiva cosciente viene concepita come «un fenomeno mentale *avanzato* o *derivato*, non come il fondamento di tutta l'intenzionalità» (Dennett 1993, p. 193).

3.1. *Funzionalismo e macchine di Turing*

Se essere in un certo stato psicologico non significa avere un particolare tipo di esperienza cosciente, quale può essere la sua natura? Un'autorevole risposta – in effetti la risposta prevalente nella filosofia della mente degli ultimi quarant'anni – è fornita dal *funzionalismo metafisico*, la tesi secondo cui uno stato mentale *è* il tipo di ente che è in virtù del ruolo causale che svolge nella vita mentale di un agente, indipendentemente dal modo in cui tale ruolo è realizzato fisicamente (Block 1980). Consideriamo la credenza di Amleto che lo zio abbia ucciso suo padre. Secondo il funzionalista, Amleto è in questo stato mentale in virtù non già della sua particolare costituzione neurobiologica, ma dello schema astratto di relazioni causali che lo stato intrattiene con stimoli sensoriali, risposte comportamentali e altri stati psicologici. La credenza è cioè causata da un'esperienza percettiva (ad Amleto appare lo spettro del padre che gli rivela di essere stato avvelenato dal fratello Claudio); ed è a sua volta causa, unitamente al desiderio di vendetta, del suo brutale comportamento nei confronti della regina Gertrude.

Un modo in cui è possibile formalizzare il funzionalismo si basa sulla teoria della computazione sviluppata dal matematico inglese Alan Turing. Negli anni Trenta questo studioso mise a punto un modello formale dell'attività di un agente che esegue un calcolo di tipo *algoritmico*, vale a dire un procedimento in grado di assolvere a un certo compito (ad esempio, le operazioni aritmetiche) applicando in modo meccanico, un numero finito di volte, un insieme di istruzioni. Tale modello è denominato *macchina di Turing* (d'ora in poi MT). Una MT consiste in un nastro la cui lunghezza si immagina infinita, suddiviso in caselle, ciascuna delle quali può contenere un solo simbolo alla volta. I simboli che una MT può adoperare appartengono a un repertorio finito, detto l'*alfabeto* di quella MT. La macchina è inoltre dotata di un cursore, in grado di spostarsi a destra o a sinistra lungo il nastro (di una casella alla volta) e capace di leggere e scrivere simboli nelle caselle del nastro. Il comportamento di una MT

è completamente determinato dalla sua *tavola di macchina*, che specifica, per ogni possibile stato in cui si trova il cursore e ogni possibile simbolo presente sulla casella corrente, qual è lo stato successivo e qual è l'azione da compiere: spostarsi a destra, spostarsi a sinistra o sovrascrivere. Consideriamo, ad esempio, la seguente tavola (tratta da Haugeland 1985):

	1	2	3	4	5	6	7
Y	_S1	_S1	YS3	YS4	YD5	YD6	_D7
_	_S1	YD2	STOP	YD5	YS3	AS3	YD6
1	1S2	AD2	AS3	1S7	AD5	AD6	1D7
A	1S1	YD6	1S4	1S4	1D5	1D6	_D2

La prima colonna a sinistra specifica gli input che la MT può ricevere (in questo caso soltanto quattro). Nella prima riga in alto sono invece elencati i sette stati che la MT può assumere. Il suo comportamento è dunque descritto esaustivamente dalla matrice di 28 celle sopra riportata. In ogni cella (eccettuata quella in cui compare l'istruzione 'STOP' che arresta la MT), i primi due simboli descrivono l'output della MT. Quest'ultima scrive un simbolo sul nastro, sposta il cursore a destra (D) o a sinistra (S), portandosi sul simbolo nella successiva casella del nastro, il quale diviene l'input di un nuovo ciclo. L'ultimo simbolo nella cella pone la MT in un nuovo stato. Ad esempio, se la MT si trova nello stato 2 e il suo input è 'A', stampa 'Y', si sposta di una casella verso destra e assume lo stato 6.

La tavola di macchina appena presentata si riferisce a una particolare MT, detta *macchina di Turing universale*. Quest'ultima è una MT in grado di assumere come programma una descrizione opportunamente codificata di un'altra MT (cioè i suoi simboli e la sua tavola di macchina) e calcolare la funzione che *essa* calcola. Un calcolatore programmabile altro non è che una *realizzazione fisica* di una MT universale (solo parziale, però, perché la memoria del computer è finita, mentre il nastro della MT è infinito). Per tale ragione, la teoria delle MT costituisce il fondamento dell'informatica teorica.

La teoria delle MT mette a disposizione del funzionalista un modo di formalizzare le proprie definizioni degli stati psicologici. La tavola di macchina sopra presentata descrive esaustivamente lo stato di una MT in base a tre elementi: (1) l'input della MT; (2) l'output della MT dato il suo stato e quell'input; (3) lo stato successivo della MT dato lo stato attuale. Non c'è bisogno di nient'altro per definire lo stato; in particolare, non c'è bisogno di far alcun riferimento al supporto che lo realizza fisicamente. Ad esempio, lo stato 2 della MT sopra presentata altro non è che lo stato in cui la macchina, se legge 'A', scrive 'Y', si sposta a destra e assume lo stato 6; oppure, se legge '1', scrive 'A', si sposta a destra e rimane nello stesso stato e così via. Pertanto possiamo dire che lo stato 2 è definito *funzionalmente*: ciò che conta è quello che la MT *fa* quando si trova in tale stato, non la sua costituzione materiale.

3.2. *Il meccanismo della causazione intenzionale*

Quale relazione sussiste tra gli stati funzionali/computazionali teorizzati dagli psicologi cognitivi e gli stati mentali postulati dalla psicologia del senso comune? Per rispondere a tale domanda può essere utile prendere le mosse da questo passo di Daniel Dennett:

> Una volta detto che una persona ha una sensazione di dolore, la localizza ed è indotta ad agire in un certo modo, abbiamo detto tutto quello che c'era da dire nell'ambito di questo vocabolario [di livello personale]. *Possiamo* richiedere ulteriori spiegazioni di come accada che una persona ritiri la mano dalla stufa rovente... [ma] se lo facciamo, dobbiamo abbandonare il livello esplicativo delle persone e delle loro sensazioni e attività per volgerci al *livello subpersonale* dei cervelli e degli eventi nel sistema nervoso. Tuttavia, una volta abbandonato il livello personale, abbiamo abbandonato anche [...] l'argomento dei dolori [...] poiché la nostra analisi alternativa non può essere affatto un'analisi del dolore, ma piuttosto di qualco-

s'altro: del movimento dei corpi umani o dell'organizzazione del sistema nervoso (Dennett 1969, pp. 93-94).

Per chiarire il pensiero di Dennett è utile tracciare una distinzione tra due forme di spiegazione: orizzontale e verticale (Bermúdez 2000). Nel brano appena citato la spiegazione psicologico-popolare (o di 'livello personale') viene presentata come una spiegazione *orizzontale*: un particolare evento (una persona ritrae la mano da una stufa rovente) viene spiegato alla luce di stati ed eventi antecedenti (la persona ha una sensazione di dolore, la localizza ed è indotta ad agire in un certo modo). Questa spiegazione è autonoma dal livello subpersonale in quanto fa *esclusivo* ricorso a stati ed eventi appartenenti al livello personale.

Pur essendo autonoma, la spiegazione orizzontale è perfettamente compatibile con un altro tipo di spiegazione caratterizzabile come *verticale*. Le spiegazione orizzontale e quella verticale perseguono obiettivi differenti: la prima ci aiuta a cavarcela nel mondo sociale; la seconda ambisce a *legittimare* la prassi mentalistica popolare:

> Si può dire in una singola frase in che cosa la psicologia computazionale riesce così bene: *la convalida delle generalizzazioni relative agli atteggiamenti proposizionali*, specificamente dei tipi di generalizzazioni relative agli atteggiamenti proposizionali più o meno di senso comune. Dunque, ad esempio, possediamo i frammenti di una teoria della percezione, e questa chiarisce come un sistema computazionale possa giungere regolarmente a credere che *p* come conseguenza causale del fatto che è visibilmente vero che *p* [...] la funzione di una teoria computazionale è chiarire il meccanismo della causazione intenzionale, mostrare come è (nomologicamente) possibile che sistemi puramente computazionali – in effetti, puramente fisici – agiscano in base alle loro credenze e desideri (Fodor 1986, p. 422).

Dunque, secondo quanto Fodor sostiene in questo passo, le spiegazioni verticali della psicologia computazionale non rivaleggiano con le spiegazioni orizzontali della psicologia del senso comune. Piuttosto il loro scopo è spiegare come le leggi intenzionali *possono essere vere*.

Questo problema di chiarire il meccanismo della causazione intenzionale, di mostrare come è possibile che sistemi puramente fisici agiscano in base a credenze e desideri, ha almeno due dimensioni.

Il primo problema riguarda gli *stati* intenzionali. Questi ultimi sono *valutabili semanticamente*: rappresentano il mondo e lo fanno in modo corretto o scorretto. Ad esempio, la mia credenza che Roma sia a sud di Torino rappresenta il mondo in un certo modo, e dunque la mia credenza è vera se e solo se il mondo è in quel certo modo – tale stato di cose è la *condizione di verità* della mia credenza. I desideri rappresentano il mondo come colui che desidera vorrebbe che fosse: il mio desiderio di avere una pizza per cena viene soddisfatto se e solo se avrò effettivamente una pizza per cena – tale stato di cose è la condizione di *soddisfacimento* del mio desiderio. Oltre a essere valutabili semanticamente, gli stati intenzionali sono *efficaci causalmente*. Il lettore rammenterà l'esempio fatto sopra: è perché Amleto crede e desidera certe cose che agisce in un certo modo. Tuttavia, si noti: valutabilità semantica ed efficacia causale non sembrano ricorrere insieme altrove. Le *proposizioni*, come queste vengono intese dai logici, sono valutabili semanticamente, ma sono oggetti astratti e non hanno efficacia causale; i tavoli e le sedie sono oggetti concreti e hanno efficacia causale, ma non sono valutabili semanticamente. Nel corso della storia della filosofia la constatazione di questa peculiarità ha fatto sorgere molti dubbi circa la possibilità di dar conto degli stati intenzionali nel quadro di un'ontologia fisicalista. Agli occhi di molti essi restano ancor oggi *creature delle tenebre*, per citare una famosa espressione di Quine.

Il secondo problema riguarda i *processi* intenzionali, vale a dire i processi causali in cui sono coinvolti stati intenzionali. Per ritornare a un altro esempio fatto in precedenza, guardando fuori dalla finestra Mario si forma la credenza che fuori piove; il suo desiderio di non bagnarsi *interagisce* con la convinzione che il modo migliore per ripararsi dalla pioggia sia utilizzare un ombrello, e *causa* l'intenzione di munirsi di tale accessorio.

Il problema che i processi intenzionali pongono al fisicalista è il seguente. Da un lato le transizioni da uno stato intenzionale all'altro sembrano costituire una sequenza *causale*, un processo in cui ogni stato è una causa parziale del successivo. Dall'altro lato, dato un certo stato intenzionale, saranno le sue *proprietà semantiche* (il suo *contenuto*) a determinare quali altri stati seguiranno (*razionalmente*) da esso. In Mario si forma l'intenzione di prendere l'ombrello perché crede che fuori piova, desidera non bagnarsi e crede che il modo migliore per ripararsi dalla pioggia sia utilizzare un ombrello. Se le credenze e il desiderio di Mario avessero un altro contenuto, la sua intenzione di prendere l'ombrello non costituirebbe più la conseguenza razionale di questi stati intenzionali.

Dunque il *contenuto* degli stati intenzionali *causa* altri stati intenzionali. Tuttavia, perché il contenuto possa svolgere tale ruolo causale, deve esserci qualcosa nella catena causale che sia sensibile a esso. E che cosa se non un *interprete interno* potrebbe essere in grado di comprendere il significato di uno stato mentale? Dicendo questo, però, siamo finiti in un vicolo cieco: la *fallacia dell'homunculus*.

Secondo Dennett (1978, pp. 117 ss. della trad. it.) il pericolo di incorrere in una simile fallacia è la ragione più importante che Skinner ha adottato per bandire le spiegazioni mentalistiche dalla psicologia. Evitarla costituisce infatti un vincolo che ogni teoria della cognizione che ambisca a un minimo di plausibilità è tenuta a soddisfare. Ossia: la spiegazione di una capacità cognitiva deve evitare di postulare un 'piccolo uomo' dentro la testa (un *homunculus*, appunto) in possesso di quella stessa capacità. Si prenda, ad esempio, la percezione visiva. L'esperienza introspettiva mi induce a descriverla come un'immagine mentale del mondo esterno. Ma qui si annida una trappola. Chiediamoci: chi è che osserva l'immagine mentale? Se rispondiamo affermando che un occhio interno ha prodotto un'altra immagine mentale, avviamo un regresso all'infinito. Se, invece, riteniamo che l'occhio interno sia in grado di estrarre informazioni da un'immagine mentale senza doverne produrre una seconda, non si capisce più perché sia stata prodotta la prima. Il problema è, evidentemente, che spiegare la visione in base a un'immagine osservata da un occhio interno significa presupporre quelle stesse capacità che ci si era proposti di spiegare. Allo stesso modo, non è una spiegazione ricondurre la coerenza semantica o razionalità dei processi intenzionali all'esistenza di un *homunculus* dotato della stessa capacità che stiamo cercando di spiegare, vale a dire la capacità di comprendere i significati degli stati intenzionali. Pertanto la meccanizzazione della razionalità, ossia la dimostrazione che un meccanismo fisico può realizzare transizioni da uno stato intenzionale all'altro conservando alcune proprietà semantiche degli stati stessi, necessita di un modo per bloccare il regresso degli interpreti interni.

A partire da *The Language of Thought* (1975), uno dei testi sacri della filosofia della psicologia contemporanea, Fodor è andato elaborando e rielaborando una soluzione ai problemi dell'intenzionalità e della razionalità, che si articola in tre tesi: (1) una tesi sulla natura degli stati intenzionali, la teoria *rappresentazionale* della mente; (2) una tesi sul fondamento naturalistico dell'intenzionalità, una teoria *causal-informazionale* del significato integrata dalla nozione di *dipendenza asimmetrica*; (3) una tesi sulla natura dei processi intenzionali, la teoria *computazionale* della mente. La congiunzione di (1)-(3) è una delle glorie della filosofia contemporanea: il funzionalismo computazionale e rappresentazionale. (Il che non esclude, ovviamente, che si tratti di una teoria estremamente controversa.)

3.2.1. *La natura degli stati intenzionali* Abbiamo detto che gli stati intenzionali hanno proprietà sia semantiche che causali, e che questo connubio le distingue tanto dalle proposizioni che dai comuni oggetti materiali. Ma immaginiamo di scrivere: 'il libro è sul tavolo'. Da un lato questa iscrizione è un oggetto materiale (occupa una certa quantità di spazio fisico, ha una certa forma geometrica ecc.) che è dotato di efficacia causale (ad esempio, riflette la luce in modi che lo rendono leggibile ed esercita una – impercettibile ma in linea di principio rilevabile – attrazione gravitazionale sulla luna). Dall'al-

tro lato questa iscrizione (un *simbolo*) verte su qualcosa ed è perciò valutabile semanticamente: è vera se e solo se il libro si trova dove essa afferma che sia. Pertanto i simboli, al pari degli stati intenzionali, sono *sia* valutabili semanticamente *sia* efficaci causalmente.

Per Fodor la storia della riflessione sulla mente sia in filosofia che in psicologia consiste in una serie di tentativi di sfruttare questa analogia tra stati intenzionali e simboli in modo da derivare le proprietà semantiche e causali dei primi dalle proprietà semantiche e causali dei secondi. La sua teoria rappresentazionale della mente è l'erede più recente di questa tradizione, asserendo che gli stati intenzionali sono relazioni fra un organismo e rappresentazioni mentali concepite come i simboli di un *linguaggio del pensiero* (spesso chiamato *mentalese*). Più precisamente:

> Per ogni organismo *O*, e per ogni atteggiamento proposizionale *A* verso la proposizione *P*, esiste una relazione psicologica *R* e una rappresentazione mentale *M* tale che:
> (a) *M* significa che *P*, e
> (b) *O* ha *A se e solo se O* ha *R* con *M*.

Come le espressioni di una lingua naturale, così le rappresentazioni del mentalese hanno proprietà sia sintattiche che semantiche. Più esattamente, queste rappresentazioni sono simil-linguistiche in quanto sono costituite da parti (costituenti) e sono strutturate sintatticamente (ossia si combinano fra loro secondo una sintassi); le loro parti *atomiche* si riferiscono a cose e proprietà del mondo; sono *composizionali*, ossia il contenuto semantico di una rappresentazione *molecolare* è funzione del contenuto semantico delle sue parti atomiche insieme alla sua struttura sintattica; hanno condizioni di verità, e di conseguenza valori di verità determinati dal modo in cui è il mondo; intrattengono fra loro relazioni logiche di implicazione.

In una prima fase della sua riflessione Fodor ha proposto come argomento principale in favore dell'ipotesi del linguaggio del pensiero la constatazione che essa era presupposta dalle migliori teorie e modelli della scienza cognitiva degli an-

ni Settanta. Ad esempio, Fodor (1975) cerca di mostrare come tale ipotesi sia presupposta da alcuni modelli di elaborazione di informazioni, deputati a spiegare fenomeni cognitivi quali l'apprendimento dei concetti come formazione e conferma dell'ipotesi, e la presa di decisione in quanto rappresentazione e valutazione da parte di un agente delle conseguenze di un certo numero di alternative di azione alla luce di un insieme di preferenze.

In seguito Fodor ha impiegato l'ipotesi del linguaggio del pensiero per spiegare due proprietà che il pensiero condividerebbe con il linguaggio: la produttività e la sistematicità. Il pensiero è *produttivo*: siamo in grado di pensare infiniti pensieri, di illimitata complessità. Ad esempio, è possibile pensare che Roma è una città calda d'estate; che Roma e Bologna sono città calde d'estate; che Roma, Bologna e Palermo sono città calde d'estate, e così via. Inoltre il pensiero è *sistematico*: la capacità di pensare un pensiero è intrinsecamente connessa alla capacità di pensare altri pensieri. Ad esempio, se sono capace di pensare che Francesca ama il fioraio, devo anche essere in grado di pensare che il fioraio ama Francesca.

Secondo Fodor la produttività e la sistematicità del pensiero sono spiegabili in base all'ipotesi di un linguaggio del pensiero caratterizzato da una sintassi e una semantica combinatorie. Il pensiero è produttivo in quanto l'applicazione ricorsiva delle regole del mentalese a rappresentazioni atomiche o molecolari dà luogo a rappresentazioni molecolari di complessità via via crescente. Il pensiero è sistematico in quanto i processi mentali accedono alla struttura in costituenti delle rappresentazioni del mentalese.

In Fig. 1.5 sono rappresentate le strutture in costituenti delle frasi 'Francesca ama il fioraio' e 'il fioraio ama Francesca'. Le due frasi – e le corrispondenti rappresentazioni mentali – condividono costituenti e struttura; l'unica differenza è che i due sintagmi nominali mutano le posizioni di soggetto e oggetto nell'albero. Chiunque sia in grado di accedere a tale struttura, sarà in grado di comprendere entrambe le frasi – e di pensare le corrispondenti rappresentazioni mentali.

Un atteggiamento proposizionale è dunque una relazione

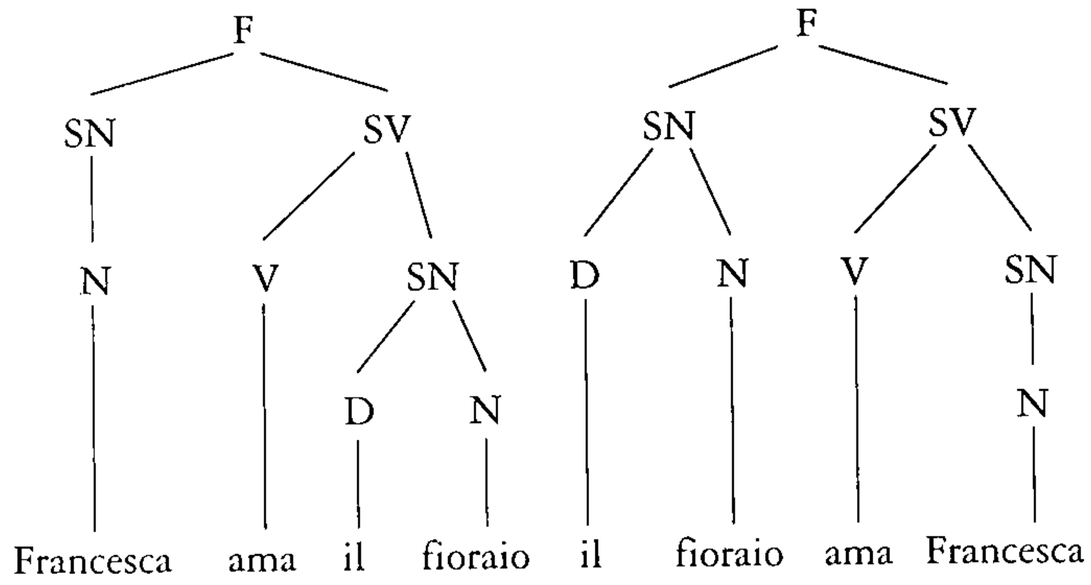


Figura 1.5. Due alberi etichettati.
Legenda: F = frase, SN = sintagma nominale, SV = sintagma verbale, D = determinatore, N = nominale, V = verbo. Le posizioni in cui compaiono le lettere maiuscole sono dette 'nodi'; le lettere maiuscole sono dette 'etichette di categoria grammaticale'; ciò che occupa un nodo è detto 'costituente'.

tra un organismo e una rappresentazione del linguaggio del pensiero (o simbolo del mentalese). L'atteggiamento eredita il proprio contenuto semantico da quello dei simboli del mentalese. Il problema dell'intenzionalità si condensa allora nel seguente interrogativo: in che modo i simboli del mentalese possono avere un significato intrinseco, cioè rappresentare un oggetto, una proprietà o uno stato di cose?

La risposta fornita da Fodor si situa nel quadro della *semantica causal-informazionale*, secondo la quale l'intenzionalità è 'naturalizzabile' mediante la riduzione alla nozione di *causa*, vale a dire una relazione oggettiva appartenente all'ordine naturale. L'idea di base è semplice: un simbolo del mentalese *significa* 'cavallo' (esprime il *concetto* di cavallo, si *riferisce* ai cavalli) se è attivato in presenza di cavalli e soltanto di cavalli, cioè se è *causato* dai cavalli e da essi soltanto. Questa idea richiede però una notevole dose di elaborazione perché, in questa versione 'grezza', va incontro a molte difficoltà, prima fra tutte il cosiddetto *problema della disgiunzione*. Sup-

poniamo che, a causa della scarsa visibilità, si scambi un asino per un cavallo (la rappresentazione CAVALLO[3], attivata dai cavalli, è attivata a volte – ad esempio, in condizioni percettive non ottimali – dagli asini). Ma allora, poiché un'occorrenza di CAVALLO può in alcuni casi essere causata da un asino, come evitare la conclusione che CAVALLO significa non già 'cavallo' bensì 'cavallo *o* asino-di-notte'?

Per risolvere il problema Fodor ha formulato la teoria del significato causal-informazionale in termini di *dipendenza causale asimmetrica*. Astraendo dai complessi dettagli tecnici, l'idea fondamentale è che le relazioni causali spurie possono stabilirsi soltanto perché sussiste la connessione causale appropriata. Io non potrei pensare 'CAVALLO!' alla vista di un asino-di-notte se non esistesse una correlazione regolare tra i pensieri che vertono sui cavalli e i cavalli. Ma l'esistenza di questa correlazione *non dipende* da una correlazione tra i pensieri che vertono sui cavalli e gli asini-di-notte. Di conseguenza le occorrenze di CAVALLO causate dagli asini-di-notte esistono soltanto perché esistono le occorrenze di CAVALLO causate dai cavalli, ma non viceversa. Pertanto, conclude Fodor, CAVALLO significa 'cavallo' e non 'cavallo *o* asino-di-notte'.

3.2.2. *La natura dei processi intenzionali* Al fine di spiegare come un meccanismo fisico possa realizzare sequenze causali di stati intenzionali che hanno luogo non già in modo arbitrario ma tenendo conto delle loro proprietà semantiche, Fodor integra la teoria rappresentazionale della mente con la teoria computazionale della mente. Quest'ultima è una tesi sulla natura dei processi intenzionali che assimila la mente a un calcolatore digitale, vale a dire a un meccanismo che immagazzina rappresentazioni simboliche e le manipola in conformità a regole *formali* (basate sulla sintassi dei simboli). Da questo punto di vista i processi intenzionali sono sequenze causali guidate dalle proprietà sintattiche ma non da quel-

[3] Qui e altrove uso il maiuscoletto per segnalare il riferimento a una rappresentazione mentale.

le semantiche dei simboli. Già Hume aveva intuito che i processi mentali sono relazioni causali tra rappresentazioni mentali; tuttavia è stato soltanto nel XX secolo, grazie ad alcuni sviluppi della logica formale e alla nascita dell'informatica teorica, che si è capito come dare consistenza a una simile intuizione senza dover ricorrere all'*associazionismo* proprio della tradizione dell'empirismo britannico.

Innanzitutto *la formalizzazione mostra come collegare la semantica alla sintassi.* La logica deduttiva può essere studiata da due prospettive differenti: la prima – la *teoria della dimostrazione* – tratta le espressioni di un linguaggio (esattamente definito) prescindendo dal significato che esse possono avere; la seconda – la *teoria dei modelli* – considera le espressioni del linguaggio dal punto di vista di ciò che è rilevante per la verità e le proprietà a questa connesse. Ora, lo scopo della teoria della dimostrazione è dar conto di importanti concetti semantici (validità, implicazione, coerenza ecc.) in termini puramente formali, senza far ricorso ad alcuna nozione semantica[4]. A tal fine il logico introduce e studia un *linguaggio formale* tenendo conto esclusivamente delle proprietà formali dei simboli quali possono essere la forma fisica o la categoria sintattica. Questo linguaggio ha un alfabeto, ad esempio:

Proposizioni atomiche: p, q, r ecc.
Connettivi: $\wedge$ (congiunzione); $\vee$ (disgiunzione);
$\rightarrow$ (implicazione materiale); $\neg$ (negazione).
Simboli ausiliari: parentesi.

Vi sono poi regole per comporre gli elementi dell'alfabeto in formule ben formate (*fbf*), vale a dire in frasi 'grammaticalmente' corrette del linguaggio. Ad esempio:

1. Se p è una proposizione atomica, allora p è una *fbf*.
2. Se p e q sono *fbf*, allora $(\neg p)$, $(p \wedge q)$, $(p \vee q)$, $(p \rightarrow q)$ sono *fbf*.
3. Non esistono altre *fbf* se non quelle ottenute iterando un numero finito di volte le prescrizioni (1) e (2).

L'applicazione ricorsiva di queste regole di composizione genera un numero infinito di *fbf*. Queste ultime possono essere manipolate in base alla loro forma o sintassi in virtù di *regole di derivazione* del tipo:

$$p, (p \rightarrow q) \vdash q$$

ossia: date due *fbf* della forma 'p' e '$(p \rightarrow q)$', è derivabile 'q'. Si noti che nel presentare questo frammento di sistema formale non si è fatto alcun ricorso a nozioni semantiche come verità o validità, né è stata fatta menzione del significato delle *fbf*. Tanto nel caso delle regole di composizione delle *fbf* che in quello delle regole che ne governano la manipolazione, la forma dei simboli è l'unico parametro di cui si è tenuto conto.

Ora, ed è questo il punto cruciale, Kurt Gödel (1930) ha dimostrato con il suo teorema di completezza per il calcolo dei predicati del primo ordine[5], che un certo insieme di regole specificabile *sintatticamente* è in grado di catturare tutte le *argomentazioni logicamente valide* del primo ordine. In altri termini, Gödel ha dimostrato che mediante assiomi e regole di derivazione si possono 'ricostruire' tutti, e solo, gli *schemi di argomentazione valida* (ossia forme di argomentazione in cui se le premesse sono vere, la conclusione non può essere falsa). Talvolta per descrivere questo fatto si dice che gli aspetti sintattici di un sistema formale 'rispecchiano' o 'imitano' i suoi aspetti semantici, e viceversa.

Per riassumere, la formalizzazione mostra che, almeno ri-

[4] L'idea della formalizzazione nasce nel contesto del dibattito sui fondamenti della matematica, motivata dall'esigenza di sganciare il problema della dimostrazione delle verità matematiche dalle intuizioni semantiche del matematico. In questo quadro, Peano, Frege, Hilbert, Whitehead e Russell riuscirono a dimostrare che parti importanti della matematica erano formalizzabili, ossia riproducibili impiegando esclusivamente definizioni, assiomi e regole formali.

[5] La logica dei predicati è un'estensione della logica proposizionale che consente di rappresentare la struttura interna delle proposizioni atomiche. Per un esempio, vedi *infra*, alla conclusione di questo paragrafo.

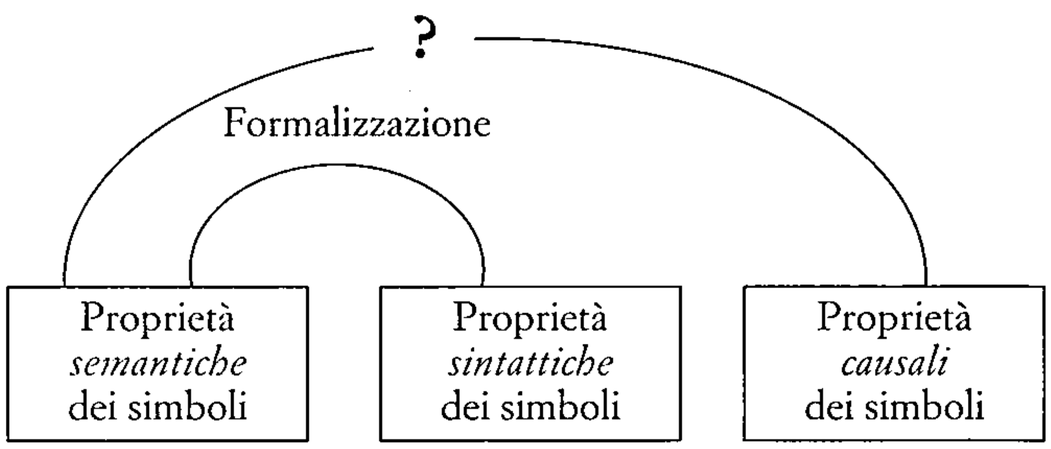


Figura 1.6. La formalizzazione mostra come collegare la semantica alla sintassi (da Horst 1996).

spetto alle argomentazioni esprimibili nel linguaggio della logica del primo ordine, è possibile specificare un insieme di simboli, di assiomi e di regole di derivazione formale tale che queste ultime, pur essendo sensibili esclusivamente alla forma sintattica dei simboli, permettono di effettuare tutte e solo le deduzioni semanticamente valide. Si ottiene così il motto formalista: «Se ti occuperai della sintassi, la semantica si prenderà cura di se stessa» (Haugeland 1985, p. 101 della trad. it.).

Inizia a delinearsi una strategia per colmare quello scarto tra la semantica e l'efficacia causale che sopra abbiamo visto impedire la meccanizzazione della razionalità: dato il nesso istituito dalla formalizzazione tra semantica e sintassi, qualora fosse possibile istituire un nesso anche tra la sintassi e l'efficacia causale, si sarebbe in condizione di collegare la semantica con la causazione in virtù della sintassi (Fig. 1.6).

Il ragionamento deduttivo è tradizionalmente considerato il paradigma di un processo razionale. Tuttavia l'intelligenza umana di certo non si esaurisce nella capacità di fare inferenze che conservano la verità. Talvolta, ad esempio, si formulano giudizi su ciò che è *probabile* sia vero. Questo è il dominio della logica *induttiva*, che si propone di stabilire regole formali, simili ai procedimenti della teoria della dimostrazione, in grado di condurre da proposizioni che sono ve-

re a proposizioni solo probabilmente vere. La possibilità di formalizzare le inferenze non deduttive è una questione assai controversa. Tuttavia, come osservano Fodor e Pylyshyn:

> Non sarebbe irragionevole descrivere la scienza cognitiva classica come un tentativo su ampia scala di applicare i metodi della teoria della dimostrazione alla modellistica del pensiero (e, analogamente, a ogni altro processo mentale che sia plausibilmente caratterizzabile in termini inferenziali: soprattutto l'apprendimento e la percezione). La teorizzazione classica si fonda sulla speranza che si possano costruire controparti sintattiche delle inferenze non dimostrative (quelle del ragionamento informale di senso comune) traendo ispirazione dal modo in cui la teoria della dimostrazione ha offerto controparti sintattiche della validità (Fodor e Pylyshyn 1988, pp. 112-113).

Supponiamo allora che si riesca a formalizzare non solo il ragionamento deduttivo, ma anche il ragionamento induttivo, abduttivo[6] e pratico. Questo risultato farebbe compiere un grosso passo avanti verso la risoluzione del problema della meccanizzazione della razionalità nel momento in cui venisse a saldarsi con la nozione di macchina di Turing. Quest'ultima, infatti, mostra *come collegare le proprietà sintattiche dei simboli alla loro efficacia causale*.

Ogni processo formalizzabile è caratterizzabile in termini di funzioni *computabili* (vale a dire funzioni descrivibili in modo algoritmico). Secondo la cosiddetta *tesi di Church-Turing*, una macchina di Turing (assumendo che nastro e tempo siano infiniti) può calcolare *qualsiasi* funzione computabile. Di conseguenza, per ogni sistema formale finito, è possibile costruire un calcolatore digitale (ossia realizzare fisicamente una macchina di Turing universale) in grado di automatizzare le derivazioni del sistema stesso. Pertanto, dal momento che le proprietà sintattiche dei simboli di un sistema formale posso-

[6] L'abduzione è un tipo di inferenza con cui si cerca di ottenere come conclusione un'ipotesi che sia in grado di spiegare i fatti descritti dalle premesse.

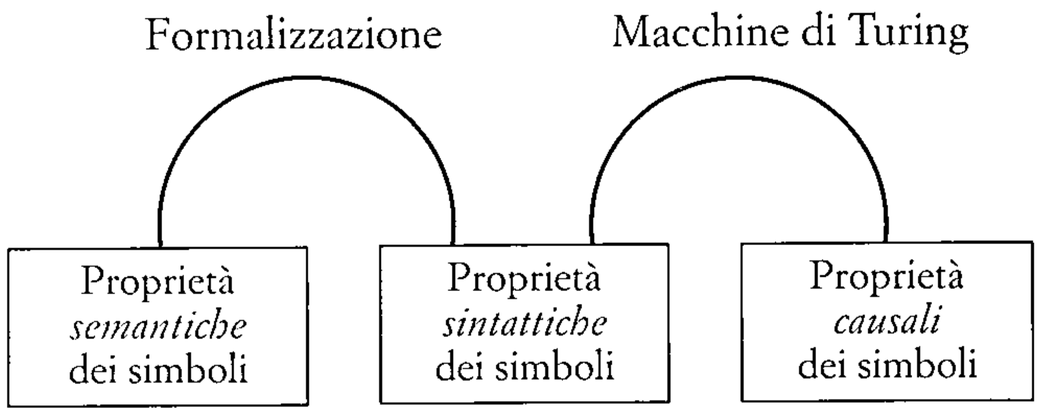


Figura 1.7. La teoria della computazione ci mostra come collegare la sintassi all'efficacia causale. Di conseguenza la congiunzione di formalizzazione e computazione getta un ponte tra causazione e semantica (da Horst 1996).

no 'imitare' le loro proprietà semantiche, e poiché è possibile realizzare fisicamente un meccanismo che manipola i simboli in modi che sono sensibili alla loro sintassi, possiamo concludere che è possibile costruire una macchina guidata dalla sintassi, le cui transizioni di stato soddisfano criteri semantici di coerenza. Questa macchina – il computer – è un modello meccanico della coerenza semantica del pensiero (Fig. 1.7).

Ora, l'ipotesi che la mente/cervello *sia* una macchina guidata dalla sintassi, le cui transizioni di stato soddisfano criteri semantici di coerenza, costituisce il fondamento del cognitivismo classico (Fodor e Pylyshyn 1988, p. 113). In quest'ottica, un processo cognitivo consiste nella manipolazione di rappresentazioni simboliche, realizzate nel sistema nervoso come le strutture dati in un calcolatore digitale. Tali rappresentazioni costituiscono un sistema formale – il linguaggio del pensiero – simile al calcolo dei predicati del primo ordine. Una regola del mentalese potrebbe allora essere:

> Per ogni rappresentazione della forma *aRb* (leggi: *a* e *b* sono legati dalla relazione *R*), è possibile sostituire $(\exists x)\ xRb$ (leggi: esiste almeno un *x* tale che *x* ha con *b* la relazione *R*).

> (Ad esempio, il pensiero MARIO AMA MARIA implica il pensiero QUALCUNO AMA MARIA.)

Tali regole 'imitano' o 'rispecchiano' le proprietà semantiche delle rappresentazioni del mentalese. Inoltre sono computabili da una macchina di Turing, il che vuol dire che sono realizzabili da una macchina fisica (il calcolatore artificiale o quello naturale, realizzato neurobiologicamente). Ciò mostra come sia possibile che un sistema puramente fisico possa realizzare inferenze deduttive e (si auspica) non deduttive, in grado di rispettare la semantica dei simboli senza ricorrere a un *homunculus* o a qualche altro agente immateriale. Le proprietà sintattiche sono le cause delle transizioni di stato razionali, la sintassi 'imita' la semantica e le proprietà sintattiche possono essere implementate in un sistema fisico.

Per riassumere, la mente è un meccanismo che opera sui simboli del mentalese in modi che, pur essendo sensibili esclusivamente alla loro struttura sintattica, ne conservano le proprietà semantiche, e dunque preservano la coerenza semantica (la razionalità) dei processi mentali intenzionali. È – per usare la bella espressione di Dennett (1978) – *un motore sintattico che imita un motore semantico.*

3.3. *L'architettura funzionale del senso comune*

Una volta che la teoria rappresentazionale della mente è integrata dalla teoria computazionale della mente, la relazione psicologica fra un organismo e le rappresentazioni del linguaggio del pensiero va intesa come una relazione computazionale. Ogni tipo di atteggiamento proposizionale è identificato con il ruolo funzionale/computazionale svolto dalla rappresentazione che ne costituisce il contenuto. Ad esempio, una certa rappresentazione è il contenuto di una credenza in quanto costituisce l'output dei sistemi percettivi e l'input di un sistema inferenziale che interagisce, in conformità ai dettami della teoria della decisione, con i desideri al fine di generare nuove rappresentazioni e condurre, infine, all'azione. O, equivalentemente, credere che *p* vuol dire che

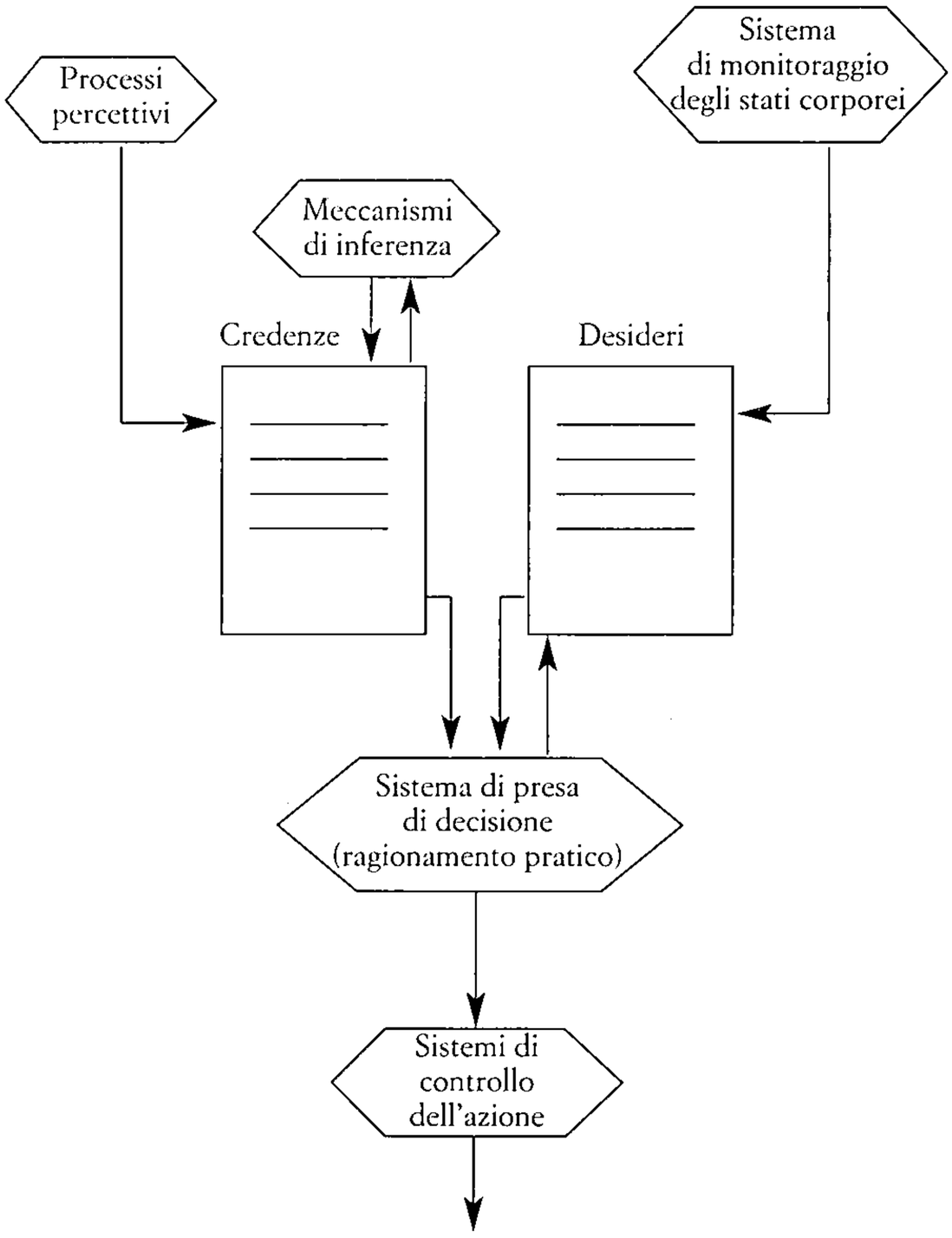


Figura 1.8. L'architettura funzionale del senso comune (da Nichols e Stich 2001).

una rappresentazione che significa *p* è disponibile a *un certo* insieme di computazioni mentali – mentre desiderare, sperare, aspettarsi che *p* significa che quella stessa rappresentazione è disponibile ad *altri* insiemi di computazioni. A tale proposito, è consuetudine immaginare, ispirandosi ai registri di un computer e ai diagrammi di flusso degli psicologi cognitivi, che le rappresentazioni del linguaggio del pensiero siano immagazzinate in una 'scatola delle credenze' o in una 'scatola dei desideri', proprio per raffigurare vividamente questa disponibilità selettiva. La Fig. 1.8 è una raffigurazione delle scatole delle credenze e dei desideri in interazione con altri meccanismi mentali.

Le credenze e i desideri in Fig. 1.8 differiscono sotto il profilo funzionale in quanto sono causati in modi diversi e hanno differenti schemi di interazione con le altre componenti della mente. Talune credenze sono causate in modo abbastanza diretto dalla percezione; altre sono derivate da credenze preesistenti attraverso processi di inferenza, deduttivi e non deduttivi. Alcuni desideri sono causati dai sistemi che monitorano vari stati corporei; altri, i 'sottoscopi', sono generati da un processo di ragionamento pratico che ha accesso a credenze e desideri preesistenti. Oltre a generare sottoscopi, il sistema di ragionamento pratico deve anche stabilire in base a quale struttura di scopi e sottoscopi agire in un qualsiasi momento. Una volta presa la decisione, questa è trasmessa a vari sistemi di controllo dell'azione, il cui compito è quello di ordinare in sequenza e coordinare i comportamenti necessari per mettere in atto la decisione.

3.4. *Una visione d'insieme*

L'idea fondamentale del funzionalismo è che l'identità di uno stato mentale è determinata dal ruolo causale che esso svolge nella vita cognitiva di un agente. In Fodor questa idea diviene una difesa sia del *realismo intenzionale*, l'idea che gli stati postulati dalla psicologia intenzionale (gli atteggiamenti proposizionali) esistono e sono causalmente coinvolti nella genesi del comportamento; sia del realismo intenzionale *scientifico*, la tesi secondo cui una psicologia cognitiva matura conterrà leggi che quantificano su stati che, al pari degli atteg-

giamenti proposizionali, sono stati rappresentazionali strutturati sintatticamente ed efficaci causalmente.

In contrasto con le teorie che caratterizzano gli atteggiamenti proposizionali come relazioni *diadiche* (soggetto-enunciato di una lingua naturale; oppure soggetto-proposizione), o negano che abbiano un carattere relazionale (concezione *monadica*), la teoria rappresentazionale della mente li definisce come relazioni *triadiche* tra un soggetto, una rappresentazione mentale e un contenuto proposizionale.

Le rappresentazioni mentali sono simboli dotati di proprietà sintattiche e semantiche. Tali rappresentazioni possono svolgere un ruolo causale specifico nella vita cognitiva di un agente in virtù delle proprietà sintattiche delle loro istanziazioni fisiche (realizzate nel sistema nervoso come le strutture dati in un calcolatore digitale). In altri termini, la sintassi determina il ruolo causale e, dunque, l'identità degli atteggiamenti proposizionali.

È possibile spiegare la produttività e la sistematicità della cognizione sulla base di una sintassi e una semantica combinatorie rese possibili dalla struttura in costituenti del sistema rappresentazionale (il linguaggio del pensiero).

La semantica delle rappresentazioni atomiche del mentalese è una teoria causal-informazionale del significato integrata dalla nozione di dipendenza asimmetrica.

Infine, secondo quella che Fodor ha chiamato *condizione di formalità*, le computazioni mentali sono processi causali sensibili esclusivamente alle proprietà sintattiche delle espressioni del mentalese, che tuttavia operano in modo semanticamente appropriato, cioè conservando la verità (teoria computazionale della mente). Ciò consente di spiegare come il pensiero possa essere nel contempo razionale e meccanico.

4. *Dire di più di quello che possiamo sapere*

Non tutti i cognitivisti hanno sottoscritto la visione conciliabilista della relazione tra psicologia computazionale e psicologia del senso comune che Fodor propone. Quest'ultimo ha

senz'altro ragione quando afferma che molte ricerche in psicologia cognitiva prendono le mosse dall'assunto della sostanziale correttezza del modo in cui il senso comune raffigura l'economia funzionale della mente (Fig. 1.8). Vi sono però anche ricerche che paiono suggerire il contrario, ossia che il nostro sistema cognitivo è organizzato secondo criteri diversi da quelli assunti dalla psicologia del senso comune. Tra questi vi sono i già citati studi sull'attribuzione causale.

Un buon esempio di questo tipo di studi è il lavoro di Latané e Darley (1970) sugli effetti esercitati dalla presenza di altre persone sul comportamento di soccorso. Grazie a una lunga serie di esperimenti, questi ricercatori hanno dimostrato che la disponibilità di un soggetto ad aiutare altri individui in difficoltà diminuisce in proporzione al numero dei testimoni presenti. Ad esempio, un gruppo di soggetti sperimentali sapeva che un individuo in un'altra stanza era colpito da una crisi epilettica: la probabilità che uno di questi soggetti si impegnasse in un comportamento di soccorso diminuiva in proporzione al numero dei soggetti componenti il gruppo. Tuttavia, una volta interrogati circa la loro condotta, nessuno dei soggetti incluse la presenza degli astanti fra le possibili cause. In luogo della causa reale, vennero menzionati vari fattori (timidezza, indecisione ecc.), tutti facilmente ricavabili da uno *stereotipo* o *modello convenzionale* dei processi che potrebbero aver luogo nella mente di un individuo coinvolto nella situazione appena descritta.

Questo esperimento è uno dei molti studi sull'attribuzione causale che hanno indotto Nisbett e Wilson (1977) a formulare la tesi secondo cui le spiegazioni offerte dai soggetti in merito alle cause del proprio comportamento sono non già ricostruzioni di stati e processi mentali reali, condotte sulla base di una consapevolezza introspettiva diretta, bensì un'attività *confabulatoria* prodotta dall'impiego di teorie esplicative socialmente condivise. Questa concezione dell'introspezione è ben espressa da Jervis:

> La nostra mente non è autotrasparente: dei suoi procedimenti non sappiamo praticamente nulla. Ciò che avviene è che registriamo

nell'autocoscienza i suoi prodotti, e sporadicamente costruiamo ipotesi su come essi nascano, chiamando queste ipotesi 'introspezione' (Jervis 1993, p. 228).

Idee analoghe sono state espresse in neuropsicologia. Consideriamo, ad esempio, le sindromi da disconnessione interemisferica. All'emisfero destro di un soggetto con 'cervello diviso' (a cui, vale a dire, sia stato reciso il corpo calloso che consente ai due emisferi cerebrali di comunicare) viene presentato un comando verbale, ad esempio 'cammina'. Il paziente di norma risponde alzandosi e avviandosi verso l'uscita. Quando gli si chiede dove vada, una tipica risposta (dell'emisfero sinistro) è 'vado di là a prendere una Coca-Cola'. È quindi ragionevole supporre che la risposta al comando sia stata mediata da un processo cognitivo inaccessibile all'introspezione e non esplicitabile verbalmente (Gazzaniga 1991).

Questo tipo di dati antintrospezionisti ha indotto Wilson (1985) a ipotizzare l'esistenza di due sistemi: il primo è un sistema inconscio a fondamento del comportamento *non verbale*; il secondo è un sistema in ampia misura cosciente, la cui funzione consiste nel tentare di verbalizzare, spiegare e comunicare ciò che avviene nel sistema inconscio. In altri termini, il sistema esplicativo verbale assume in ingresso le informazioni provenienti dal sistema inconscio e, su questa base, compie inferenze basate su repertori di razionalizzazioni consensualmente ritenute accettabili.

Riflettendo su questa ipotesi, Stich ne ha evidenziato il potenziale critico nei confronti del concetto popolare di credenza. Un principio fondamentale della psicologia del senso comune asserisce che il nostro sistema cognitivo è organizzato in modo tale che il medesimo stato che porta ad asserire in modo sincero la proposizione *p* può condurre anche a vari comportamenti non verbali. Ad esempio, lo stato che determina il mio avvicinamento al frigorifero della cucina, è lo stesso stato responsabile dell'asserzione 'credo che nel frigorifero ci sia una birra'. Tuttavia dall'ipotesi dei due sistemi discende che l'economia funzionale della mente presupposta dalla psicologia del senso comune è radicalmente sbagliata. E in quei casi in cui il sistema esplicativo verbale ci induce a dire *p* e il sistema inconscio ci induce a comportarci come se credessimo un'altra proposizione, non vi è modo di stabilire a quale delle due proposizioni crediamo. In questo quadro, suggerisce Stich, si può cominciare a nutrire qualche dubbio circa la tenuta teorica del concetto di credenza. Quanto meno la plausibilità della concezione dell'architettura funzionale della mente propria del senso comune, la sua idoneità a costituire una base per edificare la teoria scientifica della mente è – con le parole di Stich (1983, p. 230) – «in larga misura un problema empirico aperto».

2
La spiegazione psicologica e la struttura della mente

Il programma computazionale in psicologia è inscindibilmente legato a una concezione *modulare* dell'architettura cognitiva, secondo la quale la mente non è – come pure siamo propensi a credere alla luce dell'introspezione – un'entità monadica, bensì un insieme di sottosistemi distinti, che eseguono funzioni molto specifiche indipendentemente gli uni dagli altri. Questa concezione dell'organizzazione della mente è il frutto di una prassi esplicativa *meccanicista* che apparenta la psicologia cognitiva alla biologia e alla neuroscienza.

Introdotta in scienza cognitiva da Noam Chomsky e David Marr, l'ipotesi della modularità della mente è divenuta enormemente influente grazie a Fodor (1983). Tuttavia, malgrado il largo consenso che questa idea riscuote, forti divergenze si appuntano sul senso e la misura in cui la mente è modulare.

1. *Dalle leggi ai meccanismi*

Secondo il modello *nomologico-deduttivo* della spiegazione scientifica elaborato dal neopositivismo logico, spiegare un fenomeno naturale (*explanandum*) significa dedurlo da un *explanans* composto da leggi universali e condizioni iniziali. In accordo con questa concezione, il comportamentismo e la psicologia matematica degli anni Cinquanta hanno concepito la spiegazione psicologica come scoperta di leggi (o regolarità nomologiche) nel comportamento (Bechtel *et al.* 1998, p. 96).

Anche Fodor parla di leggi, riferendosi però alle generalizzazioni della psicologia intenzionale. A prima vista questo può apparire un uso indebito della nozione di legge. Infatti le generalizzazioni intenzionali sono *sempre* limitate da clausole 'a parità di condizioni' (in latino: *ceteris paribus*) – ad esempio:

> se *X* desidera che *p*, e *X* crede che *p* se *q*, allora, *a parità di condizioni*, *X* farà in modo di causare *q*
> (se Mario desidera rilassarsi, e crede che una commedia brillante servirà allo scopo, allora si recherà al cinema a meno che non gli capiti fra capo e collo qualche imprevisto).

E l'ineliminabilità di queste clausole sembra distinguere le generalizzazioni psicologico-popolari dalle leggi 'rigorose' da cui, almeno in linea di principio, dovrebbe essere possibile espungere le eccezioni. Ma è pur vero che leggi di questo tipo si danno solo nella fisica di base[1]; nelle leggi di tutte le altre scienze – *scienze speciali* quali ad esempio la biologia, la psicologia, l'economia, la geologia e la meteorologia – le clausole *ceteris paribus* svolgono un ruolo essenziale. Per fare un esempio (preso in prestito da Fodor 1987, pp. 4-5), si può immaginare una legge della geologia che asserisce che un fiume meandrico erode i propri argini. Questa legge vale soltanto a parità di condizioni: ad esempio, non si applica nel caso in cui sia stata costruita una diga o sia stato eretto un muro di calcestruzzo per proteggere gli argini. Sotto questo aspetto, dunque, le generalizzazioni della psicologia del senso comune possono essere considerate leggi nella stessa misura in cui lo sono le leggi delle scienze speciali.

Tuttavia, come abbiamo visto, queste leggi sono soltanto il punto di partenza per la costruzione di una seria psicologia cognitiva. Scopo di quest'ultima è infatti scoprire i *meccanismi* (le regole e le rappresentazioni della psicologia computa-

[1] Ma vedi Cartwright (1983) che fornisce ragioni per ritenere che nemmeno la fisica di base sia immune da leggi *ceteris paribus*.

zionale) che realizzano le leggi intenzionali. In tal modo, muovendo dal livello delle leggi a quello dei meccanismi, la psicologia cognitiva si inserisce nel solco di una tradizione che ambisce a spiegare i sistemi naturali ispirandosi al funzionamento delle macchine (il computer è solo l'esempio più recente). Una macchina è un sistema composto da parti che eseguono attività differenti e interagiscono reciprocamente al fine di generare il comportamento del sistema. Pertanto una spiegazione che prenda a modello il funzionamento di una macchina – una spiegazione *meccanicista* – consiste nella specificazione di un insieme di meccanismi in cui l'attività complessiva di un sistema è scomposta in varie sottofunzioni, fra loro interagenti, localizzate in determinate componenti fisiche del sistema.

1.1. *Scomposizione e localizzazione*

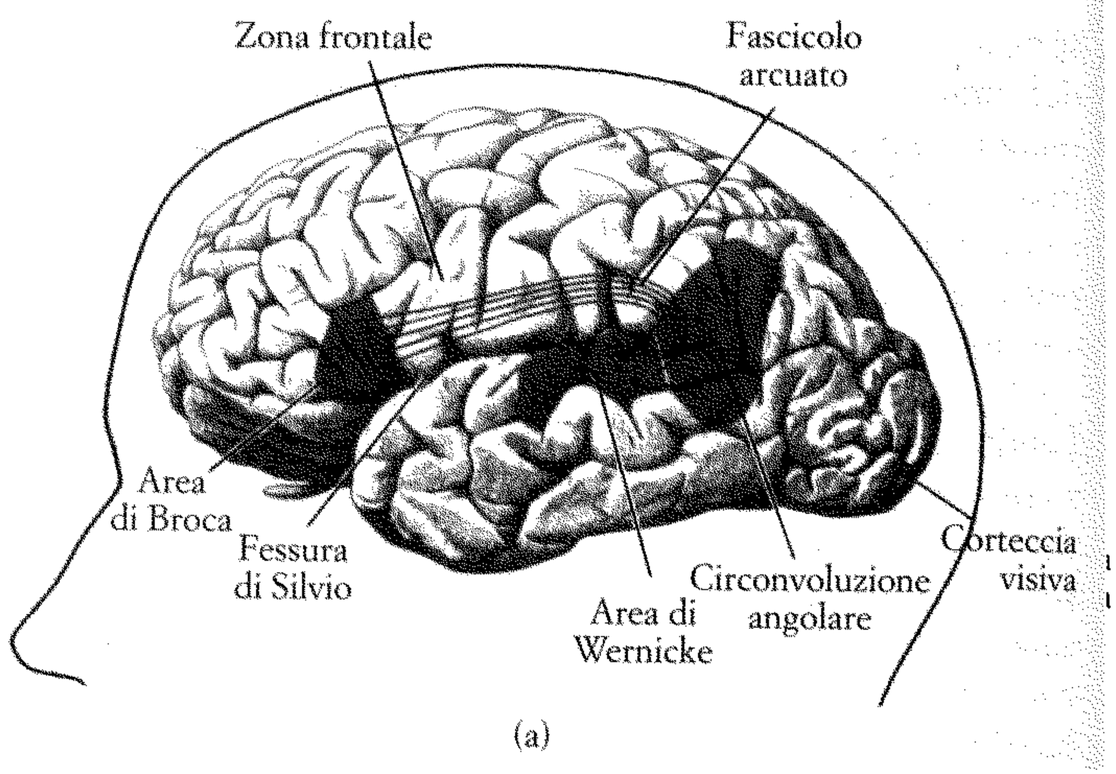
La storia delle scienze biologiche e psicologiche rivela che una spiegazione meccanicista è il frutto dell'applicazione di due strategie euristiche. La prima è la *scomposizione*, un'operazione che procede in base all'assunto secondo cui una funzione complessa eseguita da un sistema è la risultante dello svolgimento di una serie di funzioni subordinate. La seconda strategia è la *localizzazione*, un'operazione che poggia sull'assunto che il sistema contiene realmente componenti in grado di espletare tali sottofunzioni.

La localizzazione *semplice* è la forma più estrema di questo stile esplicativo: si identifica una funzione eseguita da un certo sistema complesso e se ne attribuisce la responsabilità a una *singola* componente del sistema. Ecco alcuni esempi. Alla fine del XIX secolo alcuni ricercatori proposero la localizzazione di una reazione ossidativa, la fermentazione, in un singolo enzima. Nel 1861 Paul Broca ipotizzò che un deficit di produzione del linguaggio (*afasia motoria*) fosse riconducibile a una lesione della terza circonvoluzione frontale dell'emisfero sinistro. Più di recente, O'Keefe e Nadel (1978) hanno localizzato nell'ippocampo un sistema di memoria

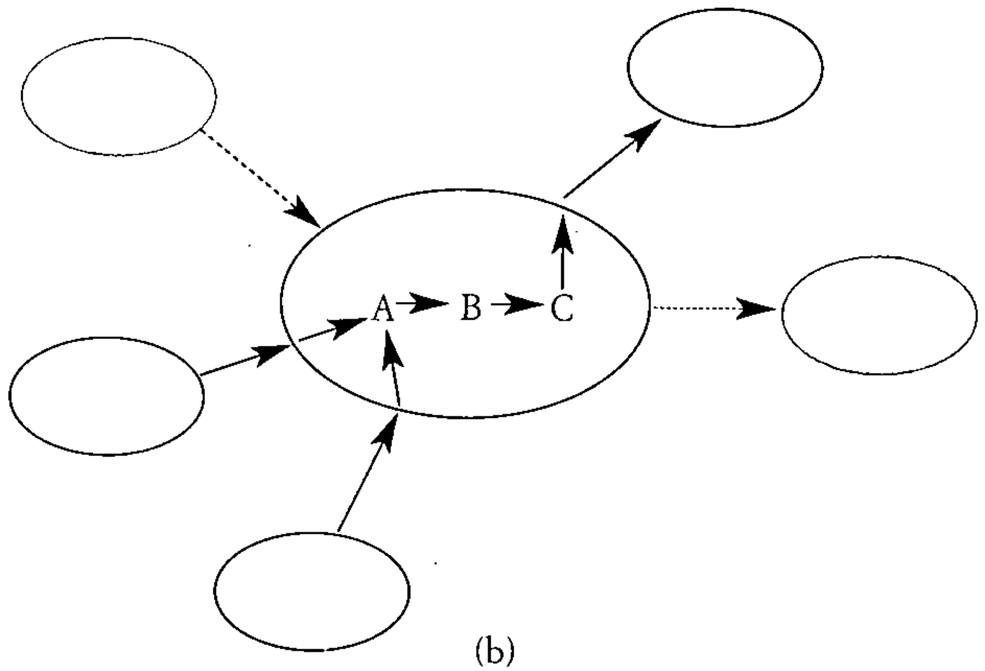
spaziale deputato all'elaborazione di una mappa oggettiva dell'ambiente. Un presupposto fondamentale di questa strategia è l'assenza di un'interazione significativa fra le componenti del sistema, il che rende quest'ultimo un mero *aggregato* delle sue funzioni indipendenti. Per avvalerci della terminologia di Herbert Simon (1969), diciamo che questo tipo di sistema è *scomponibile in modo semplice*.

Una localizzazione semplice è solo il primo passo verso l'elaborazione di una spiegazione meccanicista in quanto essa non mostra *come* un certo sistema sia in grado di espletare una determinata funzione, ma si limita a isolare un singolo sottosistema responsabile. Ciò malgrado, la messa a punto di una localizzazione semplice ha una valenza euristica: immancabilmente, infatti, emergono prove che, denunciando i difetti della localizzazione, tracciano le direttrici lungo le quali va condotta la sua revisione. Si passa così a una localizzazione *complessa*, che prevede la *scomposizione* della funzione in studio in un insieme di funzioni subordinate e la *localizzazione* di queste ultime nel sistema fisico.

Ritorniamo alla neuropsicologia dell'afasia. Dopo che Broca ebbe attribuito la funzione di produzione del linguaggio a una regione del lobo frontale, Karl Wernicke individuò una seconda regione, situata nel lobo temporale, che produceva un deficit di comprensione del linguaggio denominato *afasia corticale sensoriale*. Ciò lo condusse a formulare un modello anatomo-funzionale in cui la funzione linguistica era suddivisa in due componenti distinte, connesse fra loro, con differenti correlati anatomici: un centro di immagini acustiche verbali nella parte posteriore dell'insula e della prima circonvoluzione temporale; un centro di immagini motorie verbali nella parte inferiore del lobo frontale. Al modello di Wernicke, Lichtheim aggiunse una terza componente, il centro dei concetti collegato agli altri due centri per rendere possibile l'emissione di messaggi verbali capaci di comunicare idee. I vari tipi di afasia risultavano dalla lesione delle fibre che collegavano questi centri. Cadute nell'oblio, negli anni Sessanta queste ricerche anatomo-cliniche sono state riprese e sviluppate da Norman Geschwind. Si prenda, ad esempio, il suo adatta-



(a)



(b)

Figura 2.1. (a) Adattamento del modello di Wernicke della pronuncia di una parola scritta (da Geschwind 1979). (b) Schema di localizzazione indiretta (le ellissi e le frecce dal contorno più marcato rappresentano gli input e gli output più importanti).

mento del modello di Wernicke della pronuncia di una parola scritta. Quando si legge una parola scritta, le sensazioni vengono prima registrate dalla corteccia visiva primaria per essere poi trasmesse alla circonvoluzione angolare. Qui la forma visiva della parola viene associata alla corrispondente rappresentazione uditiva nell'area di Wernicke. Da qui, attraverso un fascio di fibre nervose denominato *fascicolo arcuato*, la rappresentazione della parola è trasmessa all'area di Broca, dove evoca un dettagliato programma di articolazione, che viene trasferito alla regione facciale della corteccia motoria. A sua volta la corteccia motoria aziona i muscoli delle labbra, della lingua, della laringe e così via (Fig. 2.1a).

Dunque, nel modello di Wernicke-Geschwind la scomposizione della funzione linguistica dà luogo a vari sottosistemi e a un'organizzazione *lineare*: ogni sottosistema esegue il proprio compito in piena autonomia, generando un output che funge da input per il sottosistema successivo. In questo caso l'organizzazione non è più aggregativa bensì *sequenziale*, con unità funzionali relativamente indipendenti. A questo proposito, sempre Simon parla di un sistema *semiscomponibile*: un certo numero di sottosistemi operano in serie lineare, ognuno realizzando autonomamente un passo dell'attività complessiva. Fatta eccezione per il fatto che i sottosistemi posteriori assumono l'input dai sottosistemi anteriori, essi sono indipendenti ed è possibile analizzare il loro contributo in isolamento (Fig. 2.1b).

1.2. *Scomposizioni ricorsive e analisi omuncolari*

Gli eredi dei modelli anatomo-clinici della neuropsicologia ottocentesca sono i diagrammi composti da 'scatole' (i processi cognitivi) e 'frecce' (il flusso di informazione) che i cognitivisti sono andati elaborando a partire dagli anni Sessanta. Qui l'approccio meccanicista diviene tutt'uno con il funzionalismo computazionale e rappresentazionale: spiegare la mente significa scomporla in un insieme di sottosistemi funzionali relativamente autonomi, in cui hanno luogo computazioni su rappresentazioni. Con le parole di David Marr: «ogni ampia computazione deve essere suddivisa in una serie

di piccoli sottoprocessi semi-indipendenti e specializzati» (Marr 1982, p. 325).

L'incontro con il computazionalismo segna una tappa fondamentale della storia delle indagini meccanicistiche sulla mente. La macchina di Turing consente infatti di rendere estremamente precisa l'idea di una scomposizione delle capacità complesse in un insieme di capacità molto semplici, realizzabili fisicamente. Consideriamo, ad esempio, una macchina di Turing che addiziona. Quest'ultima ha un alfabeto di tre simboli '1', '0', 'X'. Ogni numero *n* può essere rappresentato da una stringa di *n* '1'. La macchina addiziona trasformando sezioni di nastro come

000X111X1111X00

in sezione di nastro come

000XX1111111X00

Il segno di punteggiatura 'X' che separa 3 e 4 è stato cancellato, trasformato in un '1' e un '1' iniziale è stato trasformato in una 'X'. La capacità di addizionare è stata scomposta in capacità semplici, ossia:

- la capacità di riconoscere '1', '0', 'X';
- la capacità di cancellare '1', '0', 'X';
- la capacità di spostare il nastro a destra;
- la capacità di spostare il nastro a sinistra.

Le scomposizioni funzionali cognitiviste cercano di fare altrettanto: data una capacità cognitiva complessa, la si considera alla stregua di una funzione aritmetica come l'addizione e la si scompone in sottocomponenti.

1.2.1. *L'architettura della visione* La teoria della visione di Marr (1982) offre un esempio canonico di come una computazione mentale possa venir scomposta in una serie di sottoprocessi semi-indipendenti e specializzati. Secondo questo paradigma della scienza cognitiva, lo scopo della visione umana è il riconoscimento della forma (o geometria) di un

oggetto. A tal fine il processo visivo si snoda attraverso tre stadi di elaborazione, ognuno dei quali costruisce un certo tipo di rappresentazione. I primi due stadi corrispondono alla visione cosiddetta *primaria*, la fase in cui vengono estratte le caratteristiche elementari dallo stimolo visivo e costruite strutture visive di complessità sempre maggiore, fino ad arrivare alla discriminazione dell'oggetto dallo sfondo. Il terzo stadio di elaborazione corrisponde invece alla visione di *alto livello*, consistente nella *categorizzazione* dello stimolo, ossia nel *riconoscimento* dell'oggetto.

Il processo prende avvio dall'immagine retinica. A ogni punto (*pixel*) di questa immagine è associabile un valore numerico che codifica il valore di intensità luminosa di quel punto (registrato dai fotorecettori della retina). L'input del processo visivo è dunque una matrice di valori di luminosità, una distribuzione irregolare di zone chiare e zone scure con varie gradazioni intermedie. In questo coacervo di grigi vanno individuati i punti dove potrebbero passare i contorni di un oggetto. Indizi della probabile presenza di un contorno sono spigoli, macchie, terminazioni e strisce. Queste caratteristiche delle superfici sono individuabili essenzialmente in base al principio che là dove c'è una discontinuità molto forte nei valori di intensità luminosa – un passaggio repentino da

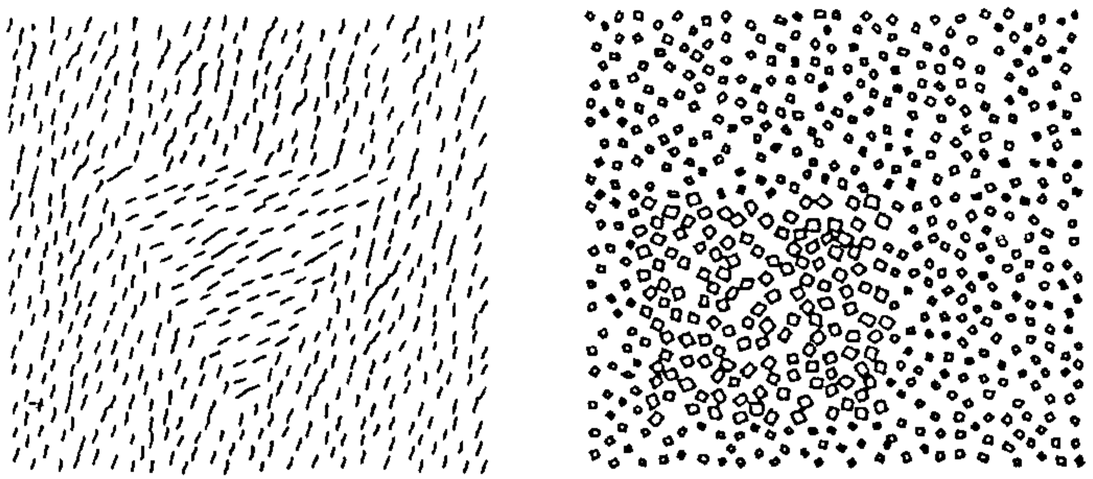


Figura 2.2 Un triangolo e un quadrato in uno schizzo primario ricco (da Sterelny 1990).

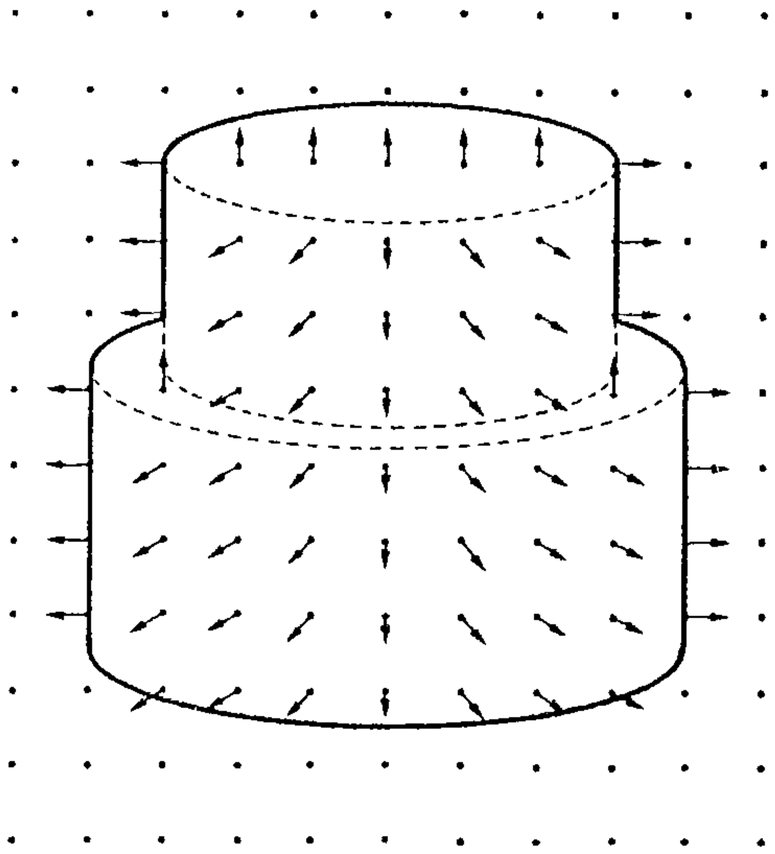


Figura 2.3. Lo schizzo 2 ½-D di due cilindri sovrapposti raffigurati come una specie di 'puntaspilli' in cui ogni spillo rappresenta la profondità e l'orientamento di una regione delle superfici dei cilindri (da Marr 1982).

una zona molto chiara a una zona molto scura (o viceversa) – è probabile che passi un contorno. Di conseguenza il primo meccanismo computazionale ipotizzato da Marr costruisce una rappresentazione esplicita di queste discontinuità, detta *schizzo primario grezzo*, in base al quale viene poi costruito lo schizzo primario *ricco*. I due schizzi differiscono in vari modi. Ad esempio, lo schizzo primario ricco rappresenta esplicitamente un'organizzazione soltanto implicita nello schizzo grezzo, raggruppando elementi simili in modo da comporre unità più piccole in unità più grandi (Fig. 2.2).

Il successivo stadio di elaborazione è la costruzione dello *schizzo 2 ½-D*, una rappresentazione intermedia fra quella bidimensionale dell'immagine a livelli di grigio e la rappresenta-

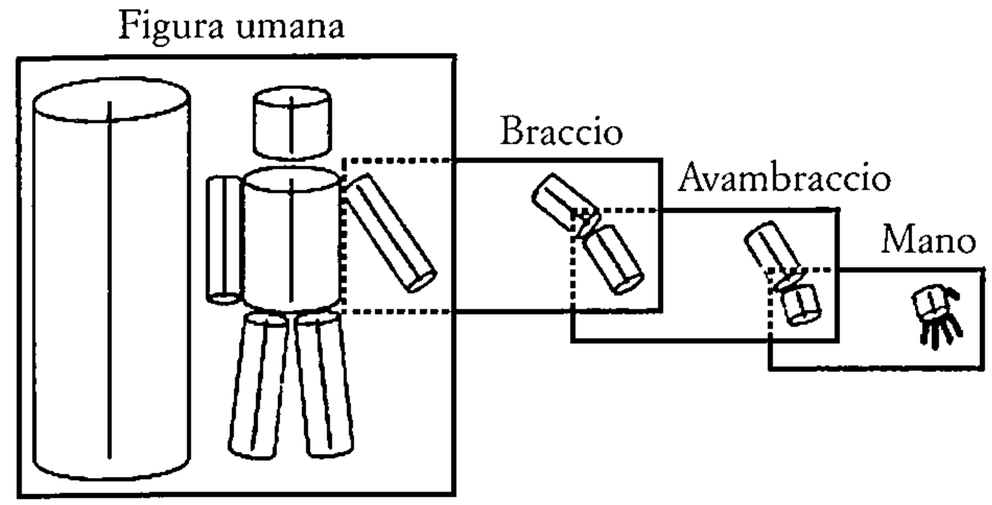


Figura 2.4. L'organizzazione gerarchica della figura umana (da Marr e Nishihara 1978).

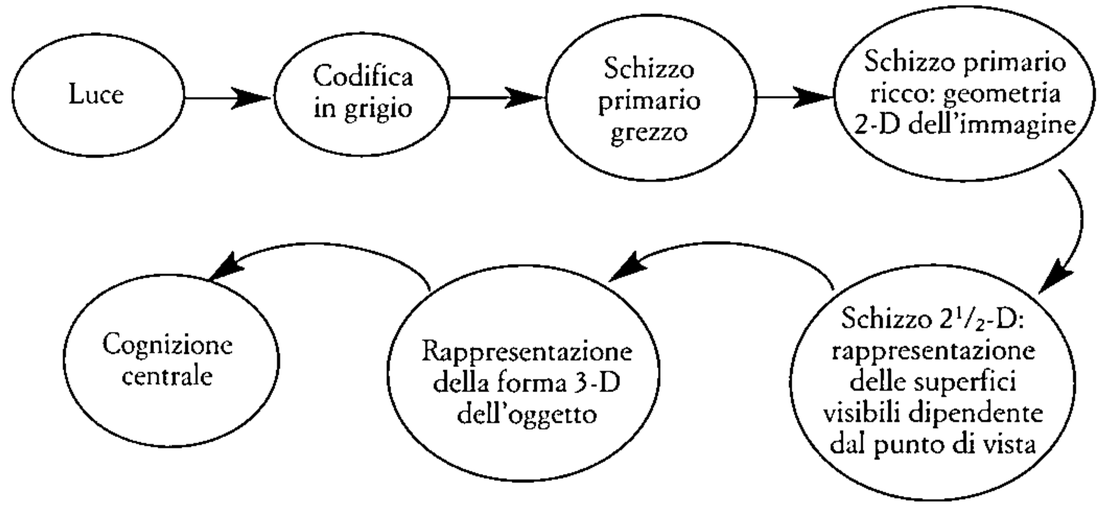


Figura 2.5. L'architettura del sistema visivo secondo Marr (da Sterelny 1990).

zione successiva che esplicita la struttura tridimensionale dell'oggetto. Si tratta essenzialmente di una rappresentazione centrata sull'osservatore, che contiene informazioni relative alla profondità e all'orientamento delle superfici ricavate da una serie di indici monoculari (ad esempio l'ombreggiatura) e binoculari (ad esempio la stereopsi) (Fig. 2.3).

Le rappresentazioni implicate nell'ultimo stadio di elabo-

razione sono tridimensionali: i *modelli 3-D*. Queste rappresentazioni sono centrate sull'oggetto anziché sull'osservatore e ciò richiede un sistema di coordinate che consenta di descrivere le relazioni spaziali tra le parti dell'oggetto rappresentato. Tali coordinate sono gli assi principali di simmetria dei cosiddetti *coni generalizzati* (cilindri e altri solidi geometrici). Un modello 3-D di un oggetto è una combinazione strutturata di questi coni (Fig. 2.4). Il riconoscimento dell'oggetto avviene confrontando tale modello con un repertorio di modelli immagazzinati in memoria.

La Fig. 2.5 offre una visione d'assieme dei vari stadi di elaborazione dell'informazione visiva secondo Marr.

1.2.2. *L'architettura della memoria* La psicologia della memoria offre un altro classico esempio di analisi funzionale. Nel 1968 Richard Atkinson e Richard Shiffrin proposero un modello della memoria che prevedeva tre sottosistemi. Innanzitutto vi era una memoria a brevissimo termine, con tempi di immagazzinamento compresi tra i 100 e i 500 msec[2], precedente anche al riconoscimento degli stimoli in ingresso. In questo tipo di memoria le informazioni erano immagazzinate in base alle loro caratteristiche fisiche: come *icone* (immagini visive) o come *echi* (immagini sonore). Questi registri sensoriali elaboravano le informazioni in parallelo e le inviavano alla memoria a breve termine, la quale manteneva temporaneamente l'informazione in uno stato di accessibilità per trasferirla, infine, nella memoria a lungo termine.

Nel 1974 Baddeley e Hitch misero in discussione la concezione della memoria a breve termine come sistema unitario, proponendo in suo luogo il concetto di *memoria di lavoro*: un sistema gerarchico deputato al mantenimento e all'elaborazione temporanea dell'informazione durante l'esecuzione di compiti cognitivi differenti (ragionamento, comprensione linguistica ecc.). Il modello della memoria di lavoro prevede un sistema attentivo, il *sistema esecutivo centrale*, che coordina un insieme di sottosistemi. Tra questi i più studiati sono il *circuito fonologico* e il *taccuino visuo-spaziale*. Il circuito fonologico è deputato al mantenimento e alla manipolazione di materiale verbale (lo utilizziamo, ad esempio, per ricordare un numero di telefono che qualcuno ci ha appena detto), oltre a essere responsabile della ricodificazione fonologica del materiale visivo (nel caso in cui, ad esempio, si deb-

[2] Per avere un'idea di quanto dura un millisecondo (msec), può essere utile sapere che pronunciare una sillaba richiede circa 200 msec, che azionare e fermare un cronometro necessita di circa 175 msec e che ogni fotogramma di un film rimane esposto per 42 msec.

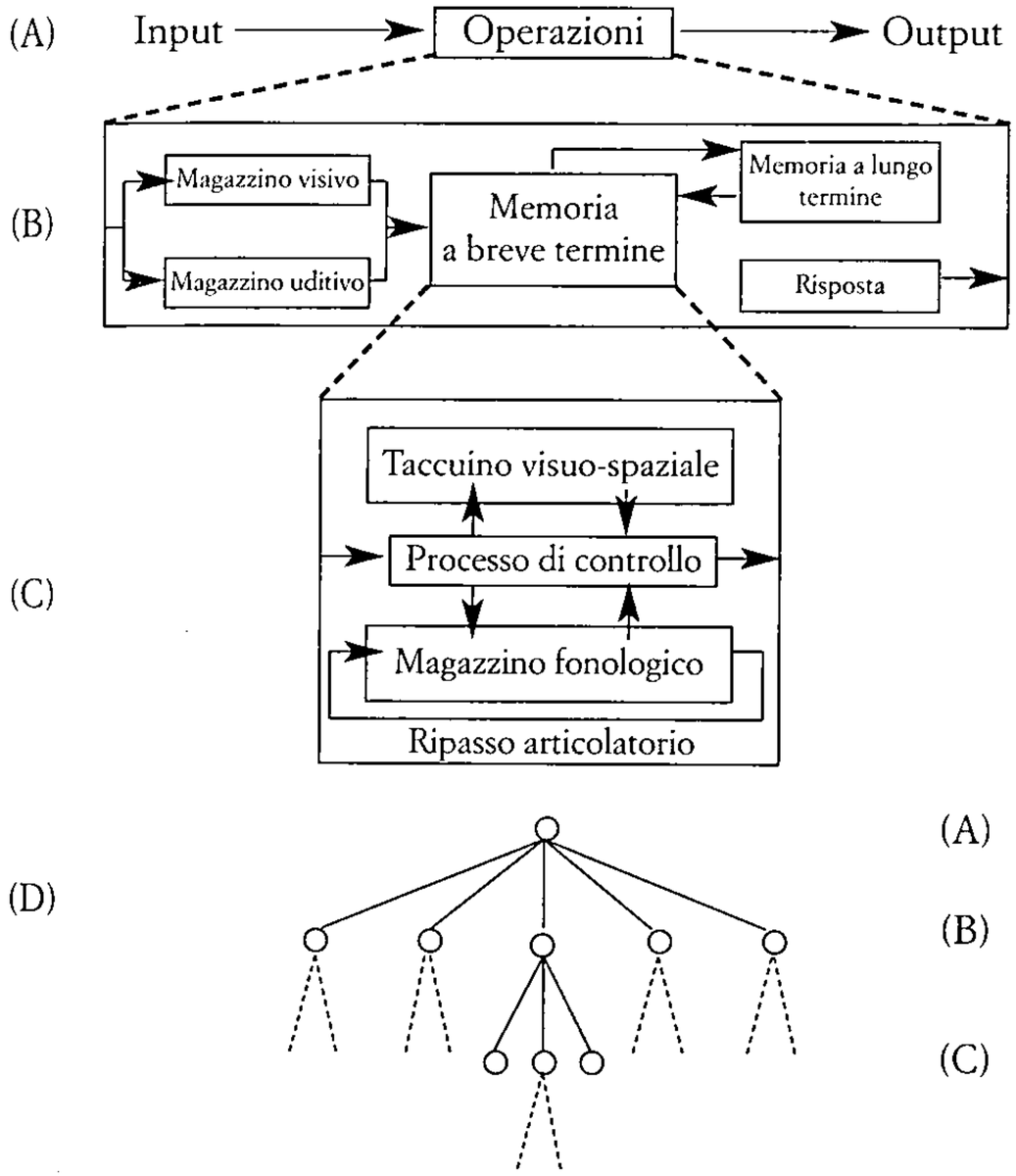


Figura 2.6. Scomposizione ricorsiva della memoria.

ba ricordare un numero telefonico letto direttamente sull'elenco anziché ascoltato). Questo sottosistema è a sua volta costituito da un *magazzino fonologico* e dal *ripasso articolatorio.* Supponiamo che qualcuno mi dica un numero di telefono: il numero, sotto forma di suono, entra nel magazzino fonologico. Se non voglio dimenticare il numero nell'arco di tempo necessario per raggiungere il telefono, lo devo 'ripassare' (mentalmente, sottovoce o ad alta voce) in modo da ravvivare la traccia di memoria.

Il taccuino visuo-spaziale è invece responsabile del mantenimento e della manipolazione di immagini mentali visive. Si ipotizza che sia anch'esso scomponibile in due sottosistemi, strettamente analoghi alle componenti del circuito fonologico: un magazzino deputato al mantenimento temporaneo di informazioni con caratteristiche visive, quali la forma o il colore di un oggetto (*visual cache*); e un meccanismo in grado di 'rinfrescare' il contenuto del magazzino stesso (*inner scribe*) (Fig. 2.6).

1.2.3. *Scaricare gli homunculi* In filosofia della mente queste scomposizioni *ricorsive* hanno ispirato una variante della dottrina funzionalista – il funzionalismo *omuncolare* – volta a neutralizzare il problema dell'*homunculus* incontrato nel cap. 1.

Il funzionalismo omuncolare integra con altre due idee la tesi secondo cui l'essenza di uno stato psicologico consiste nel suo ruolo causale. La prima è una visione *composita* della mente: quest'ultima non è un entità monadica, ma un insieme di sottosistemi funzionali che operano come unità distinte e con un grado più o meno elevato di autonomia. L'intelligenza generale e flessibile che caratterizza l'essere umano è la risultante dell'interazione fra intelligenze sempre più specializzate: «homunculi ottusi perché monomaniaci cooperano fra di loro per costituire quelli più brillanti» (Sterelny 1990, p. 13). Fuor di metafora, funzioni come la visione e la memoria vengono scomposte in funzioni via via più elementari, che interagiscono fra loro per realizzare il compito complessivo.

La seconda idea è quella che abbiamo visto all'opera nei modelli cognitivisti della visione e della memoria: la scomposizione va applicata in modo *ricorsivo*, ossia ogni *homunculus* è a sua volta visto come un complesso di *homunculi* più specializzati e dunque progressivamente più semplici. Idealmente, la scomposizione continua finché non si raggiunge un livello in cui i compiti che gli *homunculi* devono realizzare sono talmente elementari da poter essere giudicati psicologicamente primitivi. Ad esempio, nella teoria della visione di Marr un'operazione primitiva è la rilevazione dei contorni delle forme nell'immagine retinica: per spiegarla dobbiamo far ricorso non già alla psicologia ma alla neurobiologia. E a quel punto, finalmente, gli *homunculi* sono stati – per dirla con Dennett (1978, p. 81) – «scaricati».

2. *La modularità della mente*

Le parti in cui psicologi e neuropsicologi scompongono i sistemi cognitivi sono spesso chiamate 'moduli'. Questo è un senso *debole* di modularità, per cui un modulo è *qualunque* meccanismo cognitivo individuato funzionalmente (tutto ciò che avrebbe la propria 'scatola' nel diagramma di flusso di uno psicologo cognitivo). Nel pittoresco idioma del funzionalismo omuncolare, potremmo dire che un modulo è uno qualsiasi dei molteplici *homunculi* che, collettivamente, costituiscono una mente. Tuttavia in scienza cognitiva circolano usi più *forti* dei termini 'modulo' e 'modularità', ed è a questi che rivolgiamo ora l'attenzione.

2.1. *Moduli chomskiani e diacronici*

Talvolta il termine 'modulo' designa un complesso di *informazioni* o una base di *conoscenze* rappresentate internamente, che consente di spiegare una determinata capacità cognitiva (produrre e comprendere le frasi di una lingua naturale, riconoscere visivamente un oggetto, prevedere il comportamento di oggetti fisici di medie dimensioni e così via). Que-

sta base di conoscenze è *specifica a un dominio*, ossia è una struttura psicologica dedicata alla risoluzione di un numero ristretto di problemi in un dominio nettamente delimitato. Quando invece una struttura psicologica è utilizzabile in un'ampia gamma di domini differenti, viene definita *generale per dominio*.

Qualifichiamo questo tipo di modulo come 'chomskiano' in quanto la sua esemplificazione paradigmatica è offerta dalla teoria del linguaggio di Chomsky. Secondo tale teoria, la competenza linguistica di un parlante/ascoltatore consiste nella conoscenza tacita della grammatica della propria lingua, vale a dire nella rappresentazione interna inaccessibile all'introspezione di un insieme di regole sintattiche. In questa prospettiva, la comprensione di una frase necessita che a questa venga assegnata una descrizione strutturale basata su categorie astratte (sintagma, nome, verbo, anafora ecc.); e le operazioni sulle frasi richiedono che queste vengano interpretate in base a tale *struttura sintagmatica astratta*, la quale non è inferibile da proprietà superficiali dei proferimenti linguistici come, ad esempio, l'ordine lineare delle parole in una frase. Consideriamo la trasformazione di una frase dichiarativa in una frase interrogativa (Chomsky 1988):

(1) *The man is here. — Is the man here?*
(L'uomo è qui. — È qui l'uomo?)

Si possono formulare due ipotesi riguardo al modo in cui un parlante forma la frase interrogativa in (1):

I_1: il parlante segue una regola *indipendente* dalla struttura, vale a dire esamina la frase dall'inizio alla fine, parola per parola; quando trova la prima occorrenza di una classe di parole (ad esempio, un verbo come *is*), la muove all'inizio della frase.

I_2: il parlante segue un regola *dipendente* dalla struttura, vale a dire la regola è identica alla prima fatta eccezione per il fatto che il parlante sceglie la prima occorrenza di *is* che segue il *primo sintagma nominale* della frase dichiarativa.

Qual è l'ipotesi corretta? A questa domanda si può rispondere esaminando le differenti previsioni che le due ipotesi consentono di formulare in relazione a frasi simili a quella in (1). In base a I_2 è possibile prevedere la frase interrogativa corretta (le parentesi evidenziano il primo sintagma nominale della frase):

(2) [*The man* [*who is here*]] *is tall. — Is the man who is here tall?*
(L'uomo che è qui è alto. — È l'uomo che è qui alto?)

I_1 invece prevede una frase malformata:

(3) *The man who is here is tall. — *Is the man who here is tall?*

In che modo un parlante/ascoltatore arriva a 'sapere' che la trasformazione dichiarativa-interrogativa è governata dalla regola della dipendenza dalla struttura? Se il processo di acquisizione della lingua avesse un carattere induttivo, fosse cioè un processo di formazione e conferma di ipotesi, il bambino dovrebbe prima di tutto mettere alla prova I_1 in quanto *più semplice* di I_2: la regola dell'indipendenza dalla struttura si riferisce infatti all'ordine lineare delle parole nella frase e non richiede al parlante di interpretare i proferimenti linguistici in base a nozioni astratte e specificamente linguistiche come quella di sintagma nominale. La sequenza dovrebbe allora essere: il bambino accetta I_1 e poi, alla luce di dati come la frase interrogativa in (2), è costretto a rigettarla. Non è però questo ciò che si osserva: il bambino non tenta mai, neppure una volta, di applicare la regola più semplice, tanto è vero che non produce mai interrogative malformate come quella in (3). Invece impiega subito la regola più complessa della dipendenza dalla struttura.

Chomsky interpreta questi dati (e numerosi altri) come prove del fatto che il bambino non acquisisce la grammatica di una lingua in virtù di processi di apprendimento generali, applicabili in vari domini. L'acquisizione di una lingua è invece un processo di *fissazione di parametri*. Secondo

Chomsky, infatti, tutte le lingue naturali condividono una componente innata, vale a dire una serie di *principi* specificati dal linguista nella forma di una *grammatica universale*. Questi principi costituiscono uno schema di base che deve essere opportunamente integrato per ricavarne le grammatiche delle lingue particolari. L'integrazione avviene nel modo seguente. Alcuni dei principi della grammatica universale contengono *parametri binari*, punti in cui si può fare una scelta tra due alternative. Ad esempio, vi è un punto della grammatica universale che è associato a due alternative: soggetto obbligatorio *versus* soggetto opzionale. Se la lingua da acquisire è l'italiano, il bambino dovrà apprendere che l'espressione del soggetto di una frase, nella lingua materna, è opzionale. Invece, per acquisire l'inglese o il francese il bambino deve apprendere che i soggetti grammaticali sono obbligatori. Pertanto un parametro linguistico (detto *del soggetto nullo*) è fissato in modo tale da consentire l'omissione del soggetto in italiano, ma non in altre lingue. In breve, la grammatica di una lingua è il prodotto della fissazione di certi parametri, la cui natura e i cui valori possibili sono specificati dalla grammatica universale. In quest'ottica, il bambino ha bisogno dei dati empirici per selezionare i valori di questi parametri e per acquisire informazioni sul lessico: tutto il resto è innato.

Quanto finora detto è condensabile in una distinzione tra due differenti nozioni di modularità (Segal 1996). La prima nozione è *sincronica*, riguarda cioè la capacità cognitiva di un soggetto in un dato istante di tempo. Per spiegare la capacità di un parlante/ascoltatore di produrre/comprendere le frasi della propria lingua, Chomsky postula un modulo inteso come un corpo di conoscenze regolato da principi validi solo per il dominio del linguaggio. Questo modulo è essenzialmente uno *stato epistemico*, lo stato mentale di un individuo che conosce tacitamente la grammatica della propria lingua, vale a dire i principi della grammatica universale, più una scelta di valori per i parametri, più un lessico.

La seconda nozione di modularità ha invece carattere *diacronico*, concerne cioè lo *sviluppo* di una certa competenza[3]. Per spiegare lo sviluppo della competenza linguistica adulta, Chomsky postula un meccanismo di acquisizione del linguaggio che assume in ingresso i dati dell'ambiente linguistico e, attraverso un processo di selezione dei valori di alcuni parametri della grammatica universale, produce in uscita la competenza linguistica adulta. Questo è il prototipo di un modulo diacronico.

2.2. *Moduli computazionali, chomskiano-computazionali e fodoriani*

Un modulo chomskiano è un sistema innato di rappresentazioni mentali che codifica una specifica classe di informazioni. Si tratta dunque di una struttura psicologica 'inerte', che dà luogo a un comportamento solo se manipolata da meccanismi cognitivi di vario genere. Tra questi figurano anche meccanismi che possiamo definire *moduli computazionali* in quanto sono sistemi di elaborazione di informazioni specifici di dominio, ossia meccanismi in grado di computare esclusivamente un numero ristretto di input – le rappresentazioni degli oggetti e delle proprietà di un certo dominio. Ad esempio, se la fonologia costituisce un dominio, un modulo computazionale della fonologia si limiterà a effettuare analisi di input concernenti oggetti e proprietà fonologiche (Samuels 2000, p. 18).

Evidentemente i due tipi di moduli possono coesistere, visto che le informazioni manipolate da un modulo computazionale possono costituire un modulo chomskiano. Ad esempio, è possibile descrivere il meccanismo deputato all'analisi della sintassi (che gli psicolinguisti chiamano *parser*) come un modulo computazionale che impiega le conoscenze contenute in un modulo chomskiano per costruire rappresentazioni

[3] Segal intende questa distinzione in senso *pragmatico*: ciò che rende la modularità sincronica o diacronica è la prospettiva adottata, vale a dire se si è o meno interessati alle modalità di sviluppo di una certa capacità modulare.

sintattiche e semantiche di frasi. Si può ipotizzare inoltre che alcuni moduli chomskiani siano accessibili soltanto a un singolo modulo computazionale. In tal caso è naturale concepire questi due tipi di moduli come una sola unità – un insieme di algoritmi più una base di conoscenze, entrambi specifici a un dominio – che definiremo *modulo chomskiano-computazionale*. (Come vedremo nel cap. 3, un modulo di questo tipo è il 'meccanismo della teoria della mente' ipotizzato dallo psicologo dello sviluppo Alan Leslie: un meccanismo computazionale che spiega o prevede il comportamento di un agente intenzionale avvalendosi di una base innata di conoscenze specifiche al dominio psicologico, la quale non è disponibile ad altri sistemi cognitivi.)

Si noti però che la mente potrebbe contenere moduli chomskiani pur non contenendo alcun modulo computazionale. Supponiamo, ad esempio, che la mente contenga sistemi di conoscenze specifiche ai domini della fisica e della geometria. Non per questo deve anche contenere meccanismi computazionali specializzati nell'elaborazione di informazioni relative a oggetti fisici o proprietà geometriche. Può essere infatti che queste conoscenze specifiche di dominio vengano utilizzate da sistemi computazionali generali per dominio. La tesi che le cose stiano davvero così – che la mente sia effettivamente composta da algoritmi generali per dominio che si applicano a basi di conoscenze specifiche di dominio – è stata recentemente denominata da Samuels (1998) il *modello della cognizione come biblioteca*.

I moduli chomskiano-computazionali si trasformano in moduli *fodoriani* qualora vengano arricchiti da una serie di proprietà (Fodor 1983, 2000). Il modularismo fodoriano è estremamente 'austero': modulari sono infatti soltanto alcuni *sistemi periferici di input* e *output*, deputati a funzioni quali la percezione visiva e uditiva, il riconoscimento delle facce, l'elaborazione del linguaggio e il controllo motorio. Non è invece modulare la cognizione *centrale*, vale a dire l'insieme delle attività cognitive in cui si impiegano concetti, si formano credenze, si traggono inferenze e si prendono decisioni.

Secondo Fodor questi moduli periferici sono sistemi computazionali specifici di dominio e innati, che assumono in ingresso informazioni dai propri *trasduttori dedicati* per poi generare un output *superficiale*. 'Superficiale' va inteso qui essenzialmente come 'non concettuale': ad esempio, come fra breve avremo modo di vedere, il modulo dell'analisi sintattica costruisce rappresentazioni della struttura grammaticale delle frasi, ma non rappresentazioni più profonde, di livello semantico o pragmatico. Un altro esempio è il modulo che costruisce lo schizzo $2^1/_2$-D: questo output non è ancora una rappresentazione concettuale vera e propria.

I moduli fodoriani sono *isolati informativamente* dal resto del sistema cognitivo, e ciò in almeno due sensi. In primo luogo l'*accessibilità* di questi moduli è *limitata*. Ciò vuol dire che i sistemi cognitivi centrali, i quali mediano le risposte coscienti, possono accedere *esclusivamente* alle rappresentazioni fornite in uscita dai cosiddetti *macro*moduli, vale a dire moduli costituiti da *micro*moduli deputati a processi computazionali più specializzati. Pertanto le rappresentazioni *intermedie* prodotte da questi micromoduli risultano *inaccessibili* alla cognizione centrale (e dunque all'introspezione). Per fare un esempio, lo schizzo primario e lo schizzo $2^1/_2$-D, in quanto output dei micromoduli della visione primaria, non sono disponibili ai meccanismi della cognizione centrale. Questi ultimi hanno accesso soltanto all'output dal macromodulo della visione (i modelli 3-D), output che rende possibile la formazione di una credenza percettiva, accessibile all'introspezione.

L'altro senso in cui i moduli sono informativamente isolati dal resto del sistema cognitivo è questo: le operazioni di un modulo utilizzano unicamente informazioni altamente specifiche contenute al suo interno, senza mai ricorrere a conoscenze di tipo più generale disponibili al resto del sistema cognitivo. In altre parole, un modulo ha accesso esclusivamente a informazioni appartenenti alla propria *base di dati*. In questo caso si parla di *incapsulamento informativo* o, per usare un'espressione di Stich e Pylyshyn, *impenetrabilità cognitiva*. Secondo Fodor «è l'incapsulamento informativo [...] a

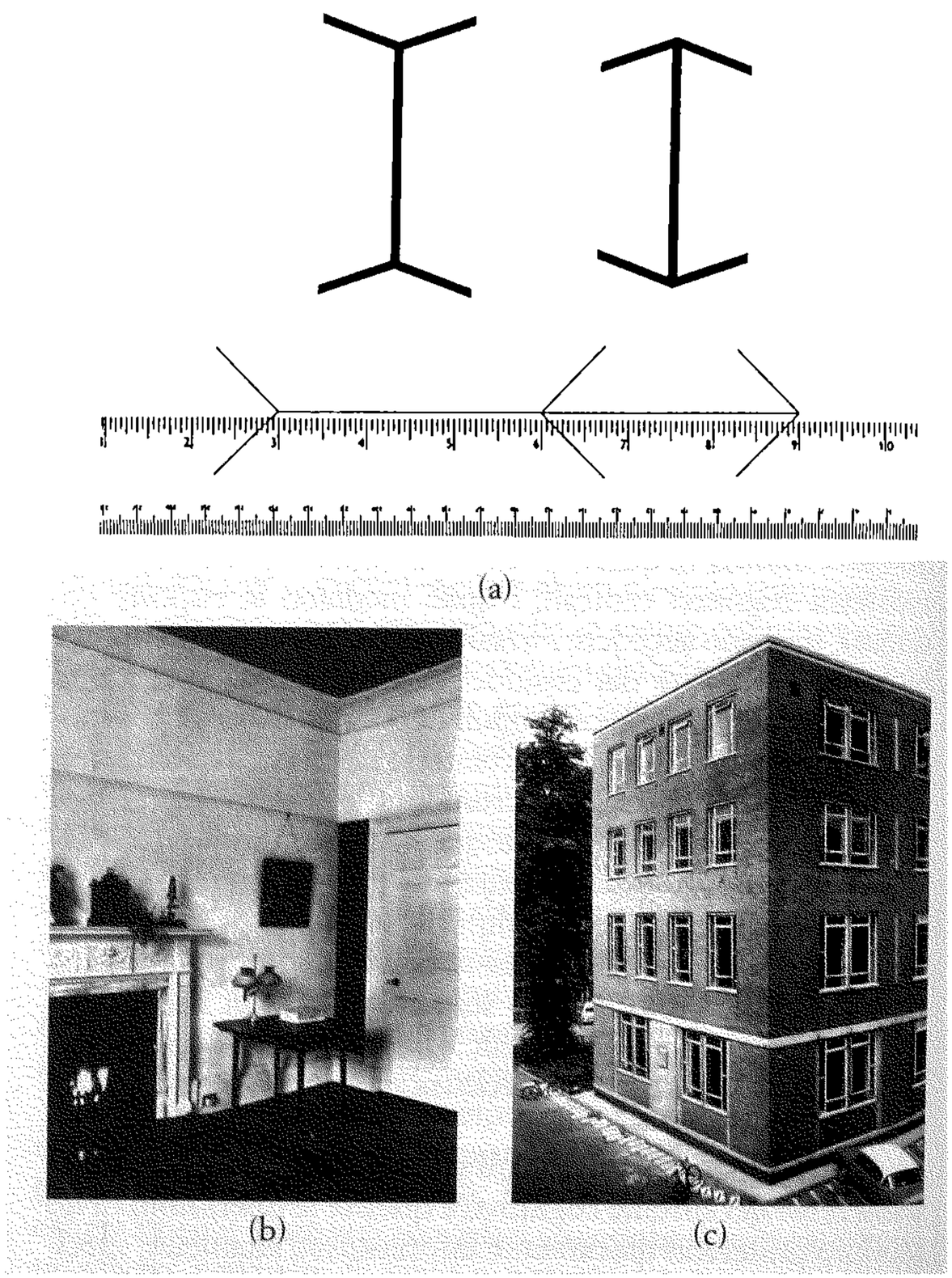


Figura 2.7. (a) L'illusione studiata da Franz Müller-Lyer nel 1889. (b) Un angolo interno: gli angoli formati dal soffitto e dal pavimento con le pareti riproducono l'illusione della freccia con i vertici piegati verso l'esterno. (c) Un angolo esterno: la linea della sommità e della base di un edificio riproduce l'illusione della freccia con i vertici piegati verso l'interno (da Gregory 1997).

costituire il cuore della modularità» (Fodor 2000, p. 79 della trad. it.). Data l'importanza del concetto, è opportuno sviscerarlo, esaminando alcune delle ragioni che sono state offerte in favore dell'ipotesi che almeno la percezione visiva, l'analisi della sintassi e il riconoscimento lessicale siano processi modulari incapsulati.

2.2.1. *Implasticità percettive* Molte illusioni percettive si oppongono all'influenza epistemica: sapere che sono illusioni non le dissolve. Esaminiamo un celebre esempio di questa persistenza: l'illusione di Müller-Lyer. In Fig. 2.7a osserviamo due frecce schematizzate, quella a sinistra con i vertici piegati verso l'esterno, quella a destra con i vertici piegati verso l'interno: benché le due frecce misurino la stessa lunghezza, la prima ci appare più lunga della seconda.

Un'accreditata spiegazione di questa illusione fa ricorso a una nostra vecchia conoscenza: la costanza di grandezza (Gregory 1997). Quando la freccia ha i vertici piegati verso l'esterno, la figura è interpretata come la proiezione tridimensionale di un angolo concavo il cui margine si allontana dall'osservatore (Fig. 2.7b). Viceversa, quando la freccia ha i vertici piegati verso l'interno, la figura è interpretata come una proiezione tridimensionale di un angolo convesso il cui margine emerge dal piano dell'immagine verso l'osservatore (Fig. 2.7c). Il risultato è che il fusto della freccia viene interpretato come più distante dall'osservatore nella figura a sinistra rispetto a quella a destra. Poiché, però, i due fusti sono di fatto della stessa lunghezza, le loro proiezioni retiniche sono identiche. Tale identità è compatibile con l'interpretazione delle figure come proiezioni piane di oggetti tridimensionali solo nel caso in cui il fusto è più lungo nella figura a sinistra rispetto a quella a destra. Infatti due oggetti posti a distanze diverse possono avere la stessa proiezione retinica solo se l'oggetto più distante è più grande. Il meccanismo della costanza di grandezza compensa l'effetto apparente della distanza, generando la distorsione percettiva (vedi cap. 1, Fig. 1.3).

Si badi: nell'ottica modularista questi assunti concernenti

le relazioni fra gli oggetti tridimensionali e le loro proiezioni bidimensionali *non* sono conoscenze ricavate dall'esperienza, che esercitano un'influsso *top-down* (dall'alto in basso) sull'analisi percettiva. Si tratta invece di *vincoli naturali*, ossia di principi di interpretazione dell'input incorporati nella base di dati innata del sistema visivo. Rispetto a ogni altra conoscenza, il sistema è impenetrabile. Ad esempio, l'illusione di Müller-Lyer si produce anche quando un'accurata misurazione ci abbia convinto che i due segmenti hanno esattamente la stessa lunghezza. La nostra ferma convinzione al riguardo non ha, a quanto pare, alcuna efficacia sul processo di visione o sul suo risultato. Semplicemente, la visione ignora tale convinzione.

2.2.2. *Analisi della sintassi e riconoscimento lessicale* Molti ricercatori concordano sul fatto che i processi di elaborazione linguistica di basso livello come l'analisi fonologica siano guidati dallo stimolo; e anche sul fatto che i processi di alto livello, come l'inferenza nell'interpretazione di una frase ironica (che dipende dalla valutazione delle intenzioni comunicative del parlante), siano invece fortemente dipendenti dal contesto, visto che si avvalgono di informazioni provenienti da svariate fonti. Sono i (molti) processi intermedi a essere oggetto di controversia; in particolare, lo sono l'analisi sintattica e i processi di riconoscimento lessicale (Fig. 2.8).

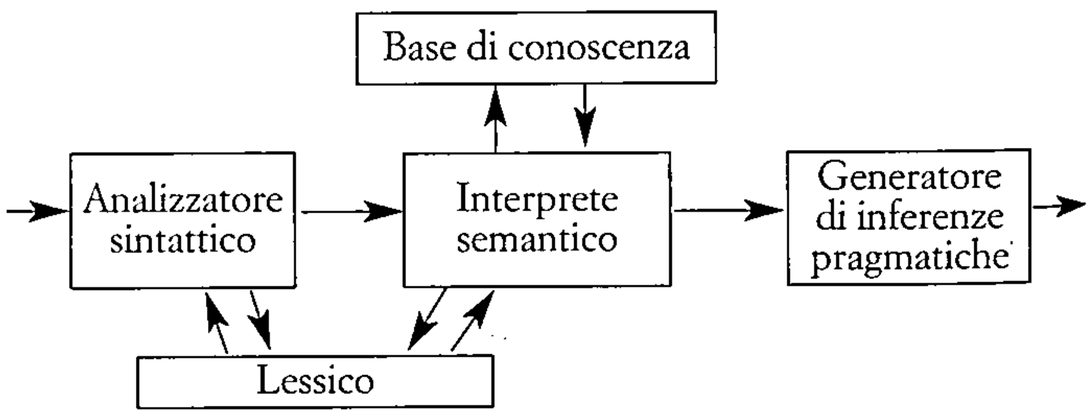


Figura 2.8. Diagramma funzionale di un generico elaboratore linguistico (da Paternoster 2002).

Analisi sintattica. A partire dalla fine degli anni Settanta il modo in cui gli esseri umani risolvono le *ambiguità sintattiche* nelle frasi è stato un importante tema di dibattito in psicolinguistica. Gran parte della controversia si è appuntata sulla questione se l'analizzatore sintattico (la componente dell'elaboratore linguistico che computa le relazioni grammaticali tra le parole) operi in modo *modulare* o *interattivo*. I modularisti sostengono che l'interpretazione iniziale della frase ricorre soltanto a indici sintattici; solo in un secondo momento entrano in gioco fattori semantici. Gli 'interazionisti' ritengono invece che esista a tutti i livelli di elaborazione un inestricabile intreccio tra sintassi e semantica.

Al problema di come l'analizzatore sintattico operi in presenza di un input linguistico ambiguo, sono state fornite essenzialmente tre risposte. Secondo la teoria del 'cammino a ritroso' (*garden path*) proposta da Lynn Frazier nel 1978, quando vengono presentate due possibili analisi sintattiche di un frammento di frase, l'analizzatore sintattico ne sceglie una e interpreta ogni ulteriore informazione nel contesto di questa analisi. Quando si trova di fronte a informazioni che non si conformano all'analisi attesa, il soggetto sperimenta una difficoltà nell'elaborazione ed è costretto ad analizzare nuovamente la frase – è obbligato cioè 'a ritornare sui propri passi'. Questa ipotesi si contrappone sia alla teoria dell'impegno minimo (*minimal commitment*) secondo cui, date due analisi possibili, l'analizzatore sintattico rinvia l'elaborazione fintantoché non riceve l'informazione che rende possibile la disambiguazione; sia alla teoria dell'*elaborazione in parallelo*, secondo cui vengono elaborate simultaneamente tutte le possibili analisi compatibili con l'input linguistico ambiguo.

La teoria dell'effetto a ritroso sostiene che l'analizzatore sintattico funziona applicando due regole che tengono conto della categoria sintattica delle parole e *non* sono guidate dall'informazione semantica o da altre informazioni di alto livello. In altre parole, per questa teoria il modulo dell'elaborazione sintattica è incapsulato.

Le due regole impiegate per scegliere la prima analisi sin-

tattica sono la chiusura tardiva (*late closure*) e l'attacco minimo (*minimal attachment*). La prima prescrive che, qualora sia grammaticalmente possibile, ogni nuovo elemento lessicale in ingresso va 'attaccato' al sintagma che è in elaborazione (ovvero il sintagma immediatamente precedente), piuttosto che essere impiegato per creare un nuovo sintagma (viene cioè rinviata la 'chiusura' del sintagma: da qui il nome). Consideriamo, ad esempio, la frase

(4) *Since John always jogs a mile seems like nothing to him.*
(Dal momento che John corre sempre un miglio non sembra nulla per lui.)

L'analizzatore sintattico inizia analizzando il sintagma 'mile' (miglio) come oggetto della frase 'John always jogs...' (John corre sempre...). Tuttavia, quando si imbatte nel verbo 'seems' (sembra) l'elaboratore si trova in difficoltà, visto che la presenza del verbo in quella posizione è incompatibile con l'analisi appena fatta. L'analizzatore sintattico deve allora rianalizzare la frase, rendendo 'mile' il soggetto di un nuovo sintagma. (Data la capacità del verbo 'seems' di discriminare fra due analisi sintattiche possibili, questa parte della frase è denominata *regione disambiguante*.)

La regola dell'attacco minimo prescrive che il nuovo materiale in ingresso vada 'attaccato' in modo da generare la più semplice struttura sintagmatica possibile. Ad esempio, la frase

(5) L'uomo colpì la donna con il bastone

è ambigua: l'uomo può aver impiegato un bastone per colpire la donna, oppure può aver colpito una donna che aveva un bastone. Queste due interpretazioni possono essere rappresentate con i diagrammi ad albero della Fig. 2.9.

La Fig. 2.9a corrisponde all'interpretazione secondo cui è l'uomo ad avere il bastone, la Fig. 2.9b rappresenta l'interpretazione secondo cui ad avere il bastone è la donna. Poiché l'analisi sintattica in (a) è più semplice rispetto a quella in (b) – l'albero di (b) ha infatti un nodo in più rispetto a quello di (a) –, la regola dell'attacco minimo predice una preferenza per interpretare 'con il bastone' come strumento del verbo

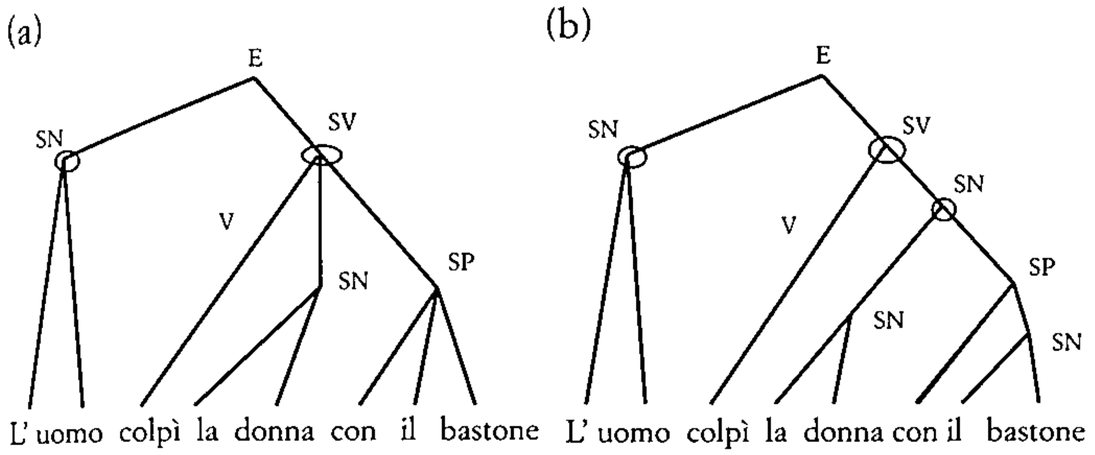


Figura 2.9. Diagrammi ad albero corrispondenti a due differenti interpretazioni della frase 'l'uomo colpì la donna con il bastone'.

('colpì con il bastone') piuttosto che come modificatore del nome ('la donna con il bastone').

Frazier e Rayner (1982) hanno fornito prove sperimentali in favore della teoria dell'effetto a ritroso. Registrando i movimenti e le fissazioni oculari di soggetti che leggevano frasi 'a ritroso' come (4), si è mostrato che l'occhio fissa più a lungo la regione disambiguante per poi ritornare indietro nella frase. Questo è esattamente ciò che ci si aspetterebbe da un soggetto che compie una prima analisi in base al principio della chiusura tardiva, nella regione di disambiguazione scopre che è errata e quindi torna indietro per effettuare una seconda analisi.

Dal momento che la teoria dell'effetto a ritroso ipotizza che l'analizzatore sintattico sia incapsulato e dunque applichi le regole senza consultare informazioni semantiche, il lettore dovrebbe sperimentare gli stessi effetti 'a ritroso' sia nel caso della frase 'l'uomo colpì la donna con i capelli rossi', sia nel caso della frase 'l'uomo colpì la donna con il bastone' – anche se, evidentemente, la prima frase si presta meglio all'interpretazione del sintagma 'con i capelli rossi' come modificatore del nome. La constatazione, invece, che l'informazione semantica influenza l'elaborazione sintattica, confuterebbe la tesi dell'incapsulamento dell'analizzatore sintattico. Siamo dunque in condizione di allestire ciò che Pylyshyn (1991)

ha chiamato *test di penetrabilità cognitiva*, vale a dire un esperimento volto a stabilire se una certa regolarità cognitiva sussiste necessariamente (ossia in virtù delle proprietà architettoniche della mente), oppure si dissolve al variare delle conoscenze e degli scopi del soggetto.

Rayner, Carlson e Frazier (1983) si sono proposti di verificare se l'informazione semantica è in grado di guidare l'analizzatore sintattico, registrando le fissazioni oculari di lettori di frasi come le seguenti:

(6) *The florist sent the flowers smiled.*
(Il fioraio dopo aver inviato – oppure: a cui erano stati inviati – i fiori sorrise.)
(7) *The actress sent the flowers smiled.*
(L'attrice dopo aver inviato – oppure: a cui erano stati inviati – i fiori sorrise.)

e manipolando il grado in cui l'informazione semantica contenuta all'inizio della frase (ad esempio, la probabilità che fiorai o attrici inviino o ricevano fiori) favoriva una particolare interpretazione. I risultati sperimentali hanno indicato che i tempi di lettura erano più lenti nel caso delle frasi che violavano le regole della chiusura tardiva e dell'attacco minimo, completamente a prescindere dalla precedente informazione semantica: una conferma, quindi, delle previsioni della teoria dell'effetto a ritroso.

Riconoscimento lessicale. Il riconoscimento lessicale è l'insieme dei processi che, dato un certo input acustico e fonetico, selezionano un elemento fra quelli contenuti nel lessico mentale. Anche qui si pone un problema per il modularista: il processo di selezione degli elementi lessicali è influenzato dalle conoscenze sintattiche e semantiche? In questo caso l'influenza *top-down* pare un dato incontrovertibile: molti dati attestano che i soggetti sperimentali identificano una parola correlata al materiale che precede una frase più rapidamente di una parola non correlata. Ad esempio, i soggetti rispondono in modo significativamente più rapido ad *ants* (formiche)

piuttosto che a *mats* (tappetini) nel contesto di una frase del tipo:

(8) *The woman was upset to find spiders, cockroaches, and...*
(La donna fu sconvolta scoprendo ragni, scarafaggi e...)

Sembra dunque legittimo congetturare che le conoscenze dei sistemi centrali (ad esempio, la conoscenza che ragni, scarafaggi e formiche sono insetti) siano inviate all'elaboratore lessicale, determinando una facilitazione del riconoscimento delle parole.

Fodor ha replicato a questo argomento tracciando una distinzione tra *associazioni lessicali* e *giudizi*. Una cosa è istituire un nesso associativo tra elementi del lessico quali 'ragno', 'scarafaggio' e 'formica'; tutt'altra cosa è disporre della conoscenza che i ragni, le blatte e le formiche sono tipi di insetti. Si può infatti immaginare che una rete lessicale, costituita da nessi associativi fissi tra elementi lessicali, faccia parte della base di dati dedicata del modulo preposto al riconoscimento lessicale. Questi nessi possono essere scambiati per giudizi, ma in realtà non lo sono, essendo invece all'origine di risposte 'ottuse', simili a riflessi:

> Qualcuno dice 'sale', tu dici 'pepe' non perché hai fatto entrare in gioco la tua credenza che il sale (nel mondo) è di norma accompagnato dal pepe; ma perché si tratta di un nesso associativo tra le forme fonologiche di questi due elementi nella tua rete lessicale (Fodor 1983, p. 81).

In quest'ottica, i differenti risultati del riconoscimento lessicale nel contesto della frase (8) sono riconducibili non già all'influenza da parte di conoscenze di alto livello, ma a una rapida attivazione automatica delle parole connesse fra loro nella rete lessicale.

Dunque tanto l'ipotesi dell'incapsulamento che quella della penetrabilità sono compatibili con i dati appena presentati. Ma supponiamo che il riconoscimento della parola 'pane' faciliti il riconoscimento della parola 'burro' a differenza, poniamo, della parola 'martello', e ciò anche quando il

contesto evidenzia che 'burro' *non* è pertinente. Ciò apparirebbe più simile a un nesso associativo 'ottuso' che a un giudizio intelligente formulato in base alle informazioni di alto livello disponibili. Un'indicazione che le cose potrebbero stare in questo modo viene da alcuni esperimenti sull'elaborazione di parole ambigue (come *coach*, *pen*, *ring*, *bug*) realizzati da Swinney (1979) e Seidenberg *et al.* (1982). I loro risultati sembrano mostrare che vi è una facilitazione automatica delle parole correlate, ad esempio, a *coach* nel senso di 'pullman' anche quando tutto nel contesto indica che solo 'allenatore' è il senso appropriato, come in (9):

(9) *The young tennis player wanted to talk to his coach.*
(Il giovane tennista voleva parlare al suo allenatore.)

Si può congetturare allora l'esistenza di un insieme iniziale di processi di riconoscimento lessicale incapsulati, che precedono il processo di disambiguazione basato su considerazioni relative al contesto e alla pertinenza. In quest'ottica, gli effetti contestuali sul riconoscimento lessicale non sono altro che il prodotto di mere associazioni.

2.2.3. *Altre proprietà dei moduli fodoriani* Oltre alla specificità di dominio, l'accessibilità limitata e l'incapsulamento informativo, altre proprietà definiscono un modulo fodoriano.

Il funzionamento di un modulo è *obbligato*: quando un modulo è in presenza del tipo specifico di input che è deputato ad analizzare, non può fare a meno di entrare in azione (ad esempio non si può evitare di udire una frase proferita nella propria lingua, né di percepire ciò su cui si posa la propria mano come la superficie di un oggetto).

L'elaborazione modulare è estremamente più *rapida* di quella dei sistemi non modulari: ciò è riconducibile all'incapsulamento (i moduli hanno bisogno di consultare soltanto la propria base di dati) e all'obbligatorietà (i moduli non perdono tempo nel decidere se elaborare o meno un input).

I moduli sono sistemi *selezionati geneticamente*, il che rende plausibile che siano realizzati in *circuiti neuronali fissi*. Da

quest'ultima proprietà e dall'incapsulamento informativo discende la *danneggiabilità selettiva*: un danno circoscritto a un gruppo di circuiti neuronali colpisce esclusivamente le funzioni eseguite dal modulo corrispondente a tali circuiti, lasciando intatte le altre funzioni del sistema.

Infine, alla luce di una serie di dati relativi all'ontogenesi delle funzioni mentali (vedi *infra*, cap. 3), è legittimo supporre che l'emergere di alcuni moduli segua un decorso evolutivo determinato in ampia misura in modo *endogeno*.

Per riassumere, un modulo fodoriano è un meccanismo computazionale specifico a un dominio che opera su una base di dati dedicata, è isolato informativamente, entra obbligatoriamente in azione, elabora l'informazione in modo estremamente rapido e produce un output superficiale che sarà successivamente elaborato da altri meccanismi. Dato il suo incapsulamento e la sua realizzazione neuronale fissa, è selettivamente danneggiabile. Il suo sviluppo, infine, è in larga misura un processo di maturazione endogena.

3. *Il problema della cognizione centrale*

Come si è già accennato, alla modularità dei sistemi periferici di input e output Fodor contrappone la non-modularità dei sistemi centrali, i quali presiedono a processi cognitivi quali la categorizzazione, il ragionamento, la risoluzione dei problemi, la fissazione delle credenze e la presa di decisione. Questo studioso insomma ci presenta la cognizione centrale come «un grande blocco olistico, in cui è impossibile ritagliare articolazioni» (Sperber 1996, p. 125 della trad. it.) (Fig. 2.10).

Qui sorge un problema profondo. Fodor ritiene che solo i processi modulari siano esplicabili con l'apparato della psicologia computazionale, e per la seguente ragione. Come abbiamo visto nel cap. 1, la psicologia computazionale classica è sottoposta al cosiddetto *vincolo di formalità*, cioè all'idea che le regole che governano le transizioni di stato di un sistema cognitivo sono sensibili esclusivamente alla *forma* delle rappresentazioni, ossia alle loro proprietà sintattiche, mentre sono

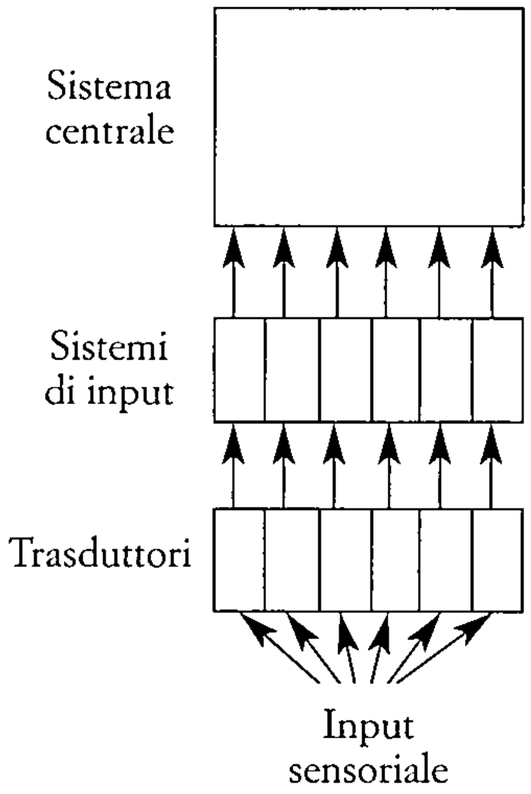


Figura 2.10. L'architettura della mente secondo Fodor.

indifferenti alle proprietà semantiche. Il lettore rammenterà che qui vale l'analogia con un calcolo logico: la forma sintattica della proposizione $p \wedge q$ si riduce all'identità e all'organizzazione dei suoi costituenti elementari; e questo è tutto quello che si deve sapere per poter inferire che la proposizione è vera se e solo se p e q sono entrambi veri – vale a dire non è necessario sapere alcunché né sul significato di p o di q, né sul mondo non linguistico. In altre parole, le proprietà sintattiche sono proprietà *locali* delle rappresentazioni, ossia sono interamente costituite dalle parti di una rappresentazione e dal modo in cui queste parti sono organizzate: «Non devi, per così dire, guardare 'fuori' da un enunciato per vedere qual è la sua struttura sintattica, non più di quanto devi guardare fuori da una parola per vedere come è scritta» (Fodor 2000, p. 20).

Ora, nel caso dei processi modulari il rispetto del vincolo di formalità è garantito dall'incapsulamento: le computazioni di un modulo incapsulato devono infatti accedere soltanto alle informazioni contenute nella propria base di dati. Il problema è che per Fodor soltanto i processi periferici hanno carattere modulare, e di conseguenza la psicologia computazionale non può essere altro che «un frammento di una psicologia cognitiva completa e soddisfacente» (Fodor 2000, p. 1). Non vi può essere una psicologia computazionale della cognizione centrale. I processi centrali hanno infatti un carattere *non locale* o *globale* od *olistico*, dal momento che possono accedere a *qualsiasi* informazione resa disponibile in *qualunque* parte del sistema cognitivo. Qui Fodor istituisce un'analogia tra la cognizione centrale e la *conferma scientifica*. Quest'ultima è *isotropica* nel senso che *ogni* conoscenza potenziale o effettiva, proveniente da qualsiasi parte del sistema di credenze, può, in talune circostanze, essere pertinente ai fini della conferma di una data ipotesi:

> Ogni cosa che lo scienziato sa è, in linea di principio, pertinente ai fini della determinazione di cos'altro deve credere. In linea di principio, la botanica vincolerebbe la nostra astronomia, se solo riuscissimo a immaginare modi in cui collegarle (Fodor 1983, p. 105).

Secondo Fodor, poiché la fissazione delle credenze è il prodotto di inferenze induttive o abduttive a partire dalle informazioni fornite dai sistemi di input e da quelle immagazzinate in memoria, anche questo processo è isotropico. Consideriamo un esempio proposto da Sterelny (1990, pp. 76-77). Supponiamo che il macromodulo della visione mi fornisca una rappresentazione dello stato del soffitto della mia camera da letto durante le prime ore del mattino. Come sappiamo, il modulo esegue questo compito accedendo a un corpo di informazioni estremamente limitato: essenzialmente, alcune informazioni concernenti la struttura della luce che impressiona la retina e alcune conoscenze generali sul mondo (ad esempio, alcuni principi di ottica geometrica). Non può invece accedere alle informazioni riguardanti il normale stato del soffitto della mia camera. Tuttavia io, al fine di stabilire se quella forma molto simile a una zanzara spiaccicata che il sistema visivo sta presentando al sistema della cognizione centrale sia veramente una zanzara spiaccicata sul sof-

fitto, e non magari un'illusione prodotta dalle prime luci dell'alba, posso accedere a una gamma praticamente infinita di conoscenze: siamo nella stagione delle zanzare? Ricordo di averne spiaccicata una sul soffitto? E così via.

Oltre a essere isotropica, la conferma scientifica è *quineiana*, nel senso che il grado di conferma che possiamo attribuire a una certa ipotesi dipende non solo dalle sue caratteristiche intrinseche, ma anche dalla sua relazione con tutte le altre credenze del sistema. In tal modo, la conferma dipende da considerazioni olistiche concernenti quelli che Fodor definisce gli «scheletri nell'armadio di Turing» (Fodor 2000, p. 84 della trad. it.): la semplicità, la centralità, il conservativismo, tutte proprietà determinate dalla struttura globale del sistema di credenze.

Anche in questo caso Fodor compie il passaggio dall'olismo della conferma all'olismo della cognizione centrale: «Qui come altrove, la psicologia cognitiva può ben rivelarsi essere la filosofia della scienza scritta in caratteri minuscoli» (Fodor 1987, p. 63). Ad esempio, al pari della conferma, la fissazione delle credenze è un processo *conservativo*. Le credenze, una volta che si sono formate, manifestano una certa *inerzia*. Di conseguenza, allorché i dati empirici premono per una revisione del sistema di credenze, le modifiche che vengono apportate sono quelle che riducono al minimo la ristrutturazione del sistema. Si tende dunque a respingere le ipotesi incoerenti con molte credenze preesistenti, specialmente quando questi impegni epistemici godono di una forte *centralità* in seno alla rete delle nostre credenze. Ad esempio, sarebbe molto difficile per me imputare l'efferato delitto di un killer seriale alla possessione demoniaca piuttosto che a qualche gravissima patologia psichica, dal momento che postulare l'esistenza di uno spirito demoniaco comporterebbe un'imponente ristrutturazione della mia ontologia. In tal modo, Fodor applica alla psicologia della credenza l'idea di Quine che i differenti costituenti di una *teoria* hanno differenti gradi di *centralità*.

Ora, per Fodor il conservativismo e la centralità sono proprietà che dipendono dalle relazioni formali, semantiche e probative fra *tutte* le credenze di un sistema. Per essere sen-

sibile a tali proprietà, un sistema computazionale dovrebbe in qualche modo passare in rassegna l'intero repertorio di credenze in modo da identificare le loro innumerevoli interconnessioni, e quindi derivare da queste caratteristiche locali una misura del conservativismo complessivo. Tuttavia, secondo Fodor, non disponiamo di alcun formalismo computazionale con cui realizzare una simile operazione. Di qui *la prima legge dell'inesistenza della scienza cognitiva*: «Quanto più globale è un processo cognitivo [...], tanto meno lo si comprende» (Fodor 1983, p. 107).

Dunque, se Fodor ha ragione nel sostenere che la maggior parte delle funzioni psicologiche sono non modulari, ci troviamo di fronte a un'enorme crepa nei fondamenti della psicologia computazionale classica.

Una possibile strategia di risposta all'argomento di Fodor circa i limiti della psicologia computazionale consiste nel capovolgere il suo ragionamento, sottoscrivendo l'idea di un legame inscindibile tra la natura computazionale della mente e l'architettura modulare, ma cercando di dimostrare che *anche* la cognizione centrale è modulare. Tale strategia assume la forma di un argomento metodologico in favore della teoria della modularità:

1. I processi cognitivi hanno una natura computazionale.
2. Tali processi devono essere locali (nel senso di avere un accesso limitato alla conoscenza di sfondo quando eseguono gli algoritmi) altrimenti non sarebbero *trattabili*, dando luogo a un'esplosione computazionale. I processi non modulari od olistici sono computazionalmente intrattabili.
3. L'unico modo di impedire a tali processi di accedere alle conoscenze di sfondo è incapsularli. Pertanto la possibilità della psicologia computazionale richiede necessariamente che la mente sia composta da una ingente quantità di meccanismi modulari.

Il problema diviene allora: è davvero possibile estendere la modularità alla cognizione centrale? Il prossimo capitolo è dedicato alla discussione di due ambiti di ricerca in cui sono stati postulati meccanismi modulari centrali.

3
Psicologia ingenua e razionalità naturale

Il tentativo di estendere la modularità alla cognizione centrale è stato denominato dall'antropologo cognitivo Dan Sperber (1994) *ipotesi della modularità massiva*. Di quest'ultima esistono almeno due versioni. La prima asserisce che l'architettura della mente è *quasi integralmente* modulare, ed è l'asse portante della cosiddetta *psicologia evoluzionistica*. La seconda versione è più moderata, ipotizzando un'architettura cognitiva in cui moduli centrali concettuali coesistono con moduli periferici di input e output e importanti sistemi o processi *non* modulari.

Nel presente capitolo discuteremo questi problemi di architettura della mente cercando di mostrare la loro centralità in due ambiti in cui l'indagine psicologica si intreccia inestricabilmente con alcune classiche questioni filosofiche: la natura della psicologia del senso comune e la relazione tra razionalità normativa e razionalità naturale.

1. *La psicologia della psicologia ingenua*

Il cap. 1 ha preso le mosse dalla teoria della teoria, la tesi della natura teorica della psicologia del senso comune. Questa idea è suscettibile di almeno due diverse interpretazioni. Innanzitutto la psicologia del senso comune è stata caratterizzata come una teoria implicita nel discorso ordinario sugli stati mentali. La classica esemplificazione di questa concezione è la *teoria causale della mente* di David K. Lewis (1972). Secondo quest'ultimo la psicologia del senso comune è una teoria *funzionalista* che determina il significato dei termini del vocabolario mentalistico nello stesso modo in cui le teorie scientifiche fissano il significato dei termini teorici, vale a dire specificandone le interconnessioni con altri termini, sia osservativi che teorici. Nella meccanica newtoniana, ad esempio, il significato di termini come 'massa' e 'forza' è definito implicitamente da leggi come 'forza = massa × accelerazione'. Analogamente, nella teoria psicologica popolare il significato dei termini del vocabolario mentalistico è definito implicitamente da generalizzazioni che – come sappiamo – specificano relazioni causali fra stimoli sensoriali e stati mentali, fra stati mentali e comportamento manifesto, fra stati mentali e altri stati mentali.

A partire dalla fine degli anni Settanta, primatologi, psicologi e filosofi hanno avviato una riflessione congiunta sulla tesi della natura teorica della psicologia del senso comune. Una volta situata nell'ambito della scienza cognitiva, la teoria della teoria ha però assunto un senso diverso da quello assegnatole da Lewis. La psicologia del senso comune viene ora intesa come una 'teoria' nel senso di un complesso di informazioni (una base di conoscenze) *rappresentate internamente*, che consente di spiegare la nostra *facoltà metapsicologica*, vale a dire la capacità di attribuire stati mentali e, in virtù di ciò, spiegare e prevedere il nostro e l'altrui comportamento.

In taluni casi una teoria intesa come struttura psicologica può essere parzialmente accessibile alla coscienza, e l'agente può esplicitare verbalmente alcune delle informazioni di cui si avvale. Il più delle volte, tuttavia, essa è *tacita*, e l'agente non ha alcun accesso introspettivo alle informazioni che guidano il suo comportamento. Abbiamo visto che questo è il caso della conoscenza della grammatica di una lingua naturale secondo Chomsky.

Supponiamo allora che la psicologia del senso comune sia una teoria di questo tipo: la teoria della mente, in larga misura opaca alla coscienza, soggiacente alla nostra facoltà metapsicologica. In tal caso le generalizzazioni psicologico-popolari a cui si riferisce Lewis potrebbero essere un'articolazione di

quel frammento della teoria della mente disponibile alla riflessione cosciente. Di conseguenza la teoria funzionalista che determina i significati dei termini mentalistici non coinciderebbe più con l'insieme delle generalizzazioni psicologico-popolari, bensì con la teoria tacita che guida effettivamente gli agenti quando pensano e parlano del mentale (Stich e Nichols 2003).

1.1. *Il test della credenza falsa*

L'indagine sulla *base naturale* della psicologia del senso comune – base che d'ora in poi chiameremo 'psicologia ingenua' – è oggi un vero e proprio settore della scienza cognitiva: la *Theory of Mind* (Teoria della Mente). Tale ambito di ricerca è andato costituendosi intorno a un paradigma sperimentale, noto come *test della credenza falsa*. Nella versione più nota di questo test, un bambino assiste a una scenetta la cui protagonista è Sally, una marionetta che mette una biglia in un cesto e quindi esce di scena. In sua assenza, un'altra marionetta, Ann, trasferisce la biglia dal cesto in una scatola. Al ritorno di Sally, al bambino viene chiesto dove Sally cercherà la biglia.

Per risolvere con successo il compito il bambino deve comprendere che Sally *crede* che la biglia si trovi ancora dove l'ha lasciata (nel cesto), ossia deve comprendere che Sally si è formata una *rappresentazione scorretta della realtà*, una credenza falsa per l'appunto. La maggior parte degli studi riportano che praticamente tutti i bambini sotto i 3 anni forniscono una risposta sbagliata, affermando che Sally cercherà la biglia nella scatola, ossia dove effettivamente si trova; intorno ai 3 anni alcuni bambini iniziano a rispondere correttamente; dopo i 4 la maggior parte dei bambini fornisce la risposta corretta, anche se alcuni studi hanno dimostrato che perfino a 5 e 6 anni si possono manifestare ancora delle difficoltà (vedi ad esempio Ruffman *et al.* 1998).

Il compito di Sally e Ann è solo uno dei molti test che in questi ultimi vent'anni sono stati impiegati per indagare la facoltà metapsicologica infantile. Questi 'compiti di teoria della mente' compongono una famiglia di paradigmi sperimentali che ha reso possibile formulare varie ipotesi circa la natura e l'ontogenesi della psicologia ingenua. Da un lato la constatazione che il bambino è capace di sofisticate prestazioni metapsicologiche assai precocemente (nei primi 3-4 anni di età) e in modo relativamente indipendente dallo sviluppo di altre capacità cognitive, ha indotto alcuni studiosi (ad esempio, Simon Baron-Cohen e Alan Leslie) a concepire tali prestazioni come il prodotto della maturazione endogena di un modulo innato (o di un insieme di moduli innati). Dall'altro lato alcuni ricercatori più legati al costruttivismo di Piaget (ad esempio, Alison Gopnik, Josef Perner e Henry Wellman) hanno sostenuto che la teoria psicologica ingenua si sviluppa nell'infanzia secondo una dinamica simile a quella delle teorie scientifiche.

1.2. *La teoria della teoria costruttivista*

La teoria della teoria è la concezione più ampiamente condivisa (e fino alla metà degli anni Ottanta, l'*unica* concezione) della psicologia ingenua. La tesi che accomuna tutte le differenti versioni della teoria della teoria è che la psicologia ingenua è un ricco complesso di informazioni rappresentate internamente a fondamento della nostra facoltà metapsicologica.

Secondo alcuni teorici della teoria, le informazioni che consentono di mentalizzare hanno la medesima struttura di una teoria *scientifica*, e vengono anche acquisite, immagazzinate e utilizzate come le teorie nel corso dell'evoluzione della scienza. Gopnik e Meltzoff (1997), ad esempio, sostengono che la psicologia ingenua possiede *tutte* le proprietà strutturali, funzionali e dinamiche che caratterizzano la maggior parte delle teorie scientifiche. Tra queste proprietà la più importante è l'*annullabilità* (*defeasibility*). Il bambino possiede una base di conoscenze innate che possono essere 'annullate', vale a dire sostituite allorché a loro sfavore si è accumulata una massa non più gestibile di *anomalie* (nel senso di Kuhn 1962). Di conseguenza, coerentemente con il costruttivismo piagetiano,

lo sviluppo cognitivo nell'infanzia e nella prima adolescenza è caratterizzato come un processo di costruzione e successivo abbandono di teorie che si susseguono una dopo l'altra – teorie che corrispondono a versioni sempre più sofisticate della teoria ingenua della mente. I meccanismi cognitivi che rendono possibile tale processo sono i medesimi meccanismi che sono alla base del mutamento nella scienza. La cognizione scientifica non è dunque altro che il protrarsi di quegli stessi processi di revisione teorica in cui i bambini sono impegnati fin dalla prima infanzia.

In taluni casi questa versione della teoria della teoria riprende da Piaget non solo l'impianto costruttivista, ma anche l'idea che lo sviluppo cognitivo è un processo causato da meccanismi *generali per dominio*, vale a dire meccanismi che danno luogo al progressivo potenziamento della capacità di risolvere problemi in qualunque dominio: fisico, biologico, psicologico, matematico e così via. Un teorico della teoria che accoglie questa concezione 'aspecifica' dello sviluppo cognitivo è Josef Perner (1991). A suo giudizio, infatti, lo sviluppo della psicologia ingenua si fonda sulla comprensione del funzionamento delle rappresentazioni *in generale*, non specificamente di quelle mentali. Innanzitutto il bambino impara a comprendere le proprietà di rappresentazioni *pubbliche* (fotografie, disegni, mappe e soprattutto proferimenti o iscrizioni linguistiche); solo in un secondo momento estende, per analogia, queste proprietà alle rappresentazioni *mentali*. In particolare, il bambino diviene capace di risolvere il test della credenza falsa quando scopre che, al pari delle rappresentazioni pubbliche, anche le rappresentazioni interne possono raffigurare stati di cose in modo scorretto.

1.3. *La teoria della teoria modularista*

Secondo la versione modularista della teoria della teoria, il ricco complesso di informazioni costituente la psicologia ingenua non ha la struttura di una teoria scientifica, essendo invece immagazzinato in uno o più moduli innati, che gradualmente si attivano nel corso dello sviluppo. Come abbiamo visto nel cap. 2, le informazioni contenute in questi moduli possono essere immagazzinate come un insieme di algoritmi dedicati (moduli computazionali); o come un insieme di conoscenze rappresentate internamente specifiche a un dominio (un modulo chomskiano); o in ambedue i modi (moduli chomskiano-computazionali). Alla seconda e alla terza di queste opzioni corrispondono due differenti versioni dell'approccio modularista alla psicologia ingenua.

1.3.1. *La psicologia ingenua come modulo chomskiano* Pur muovendosi in una prospettiva ancora costruttivista, Carey e Spelke (1996) concepiscono la psicologia ingenua come un modulo chomskiano. Queste studiose ritengono infatti che lo stato iniziale dello sviluppo infantile sia caratterizzato da un ricco repertorio di conoscenze innate e organizzate secondo domini cognitivi distinti, denominati *domini di base* (*core domains*). Un dominio di base riguarda la conoscenza del mondo fisico: gli oggetti fisici di medie dimensioni sono le principali entità di questo dominio; e il principio esplicativo più profondo è quello del *contatto*, secondo cui gli oggetti agiscono l'uno sull'altro se e solo se entrano in contatto fisico. Un altro dominio concerne la conoscenza del mondo psicologico: in questo caso, le entità del dominio sono agenti intenzionali (in modo particolare persone e animali) e i principi esplicativi più profondi sono aspetti della causalità intenzionale. A differenza di Perner, le due ricercatrici sostengono che nel dominio psicologico la nozione di rappresentazione mentale è nettamente distinta e non derivata dalla nozione di rappresentazione pubblica. In altre parole, il bambino giunge a cogliere la natura rappresentazionale delle credenze non già imparando come funzionano proferimenti e iscrizioni linguistiche, fotografie, disegni o mappe, bensì impegnandosi in un processo di graduale arricchimento di una base di conoscenze innata e *specifica al dominio psicologico*.

1.3.2. *La psicologia ingenua come modulo chomskiano-computazionale* Sappiamo però che un modulo chomskiano è una struttura psicologica inerte, che dà luogo a un comportamento solo se manipolata da qualche meccanismo cognitivo. Si tratta allora di stabilire se il modulo chomskiano della psicologia ingenua costituisca la base di dati di un sistema computazionale specifico oppure generale per dominio (vedi cap. 2, par. 2.2).

Una risposta nel senso della specificità di dominio viene dalla più sofisticata teoria della psicologia ingenua oggi disponibile: l'ipotesi del 'meccanismo della teoria della mente' (d'ora in poi ToMM, dall'inglese *Theory of Mind Mechanism*), sviluppata nel corso degli ultimi vent'anni da Alan Leslie (1987; 1994; 2000). Leslie ritiene che i sopracitati domini di base relativi al mondo fisico e a quello psicologico siano repertori di conoscenze elaborate da tre differenti moduli computazionali: ToBy, ToMM1, ToMM2. Mentre ToBy (da *Theory of Body mechanism*, 'meccanismo della teoria dei corpi fisici') ha il compito di stabilire se il movimento di un oggetto è causato da forze interne o esterne, ToMM1 e ToMM2 si occupano delle proprietà intenzionali degli agenti.

ToMM è dunque un macromodulo chomskiano-computazionale composto da due sottocomponenti. La prima contiene concetti come META o SCOPO ed è dedicata all'interpretazione *finalistica* o *teleonomica* delle azioni di un agente. La seconda include concetti quali FINZIONE, CREDENZA e DESIDERIO e inferisce a partire da una rappresentazione del comportamento di un agente, una *metarappresentazione*, ossia una rappresentazione dell'atteggiamento proposizionale che spiega il comportamento in questione. Più precisamente, una metarappresentazione è una descrizione del comportamento centrata sull'agente, costituita da una relazione triadica che esplicita quattro tipi di informazioni:

1. un agente;
2. una relazione informativa che specifica l'atteggiamento dell'agente (fingere, credere, desiderare ecc.);

3. un aspetto della realtà che àncora l'atteggiamento dell'agente;
4. il contenuto dell'atteggiamento dell'agente.

Una metarappresentazione descrive quindi un agente in quanto caratterizzato da un particolare atteggiamento nei riguardi di uno specifico contenuto e in riferimento a un determinato aspetto della realtà. Consideriamo, ad esempio, la comprensione infantile del gioco di finzione. Il bambino sembra impiegare per la prima volta l'atteggiamento proposizionale FA FINTA tra i 18 e i 24 mesi. Ciò gli consente di comprendere il comportamento della madre – intenta, ad esempio, a conversare con una banana – facendo riferimento allo stato mentale della madre, catturato dalla seguente metarappresentazione:

> La mamma FA FINTA (di) questa banana (che) 'è un telefono'.

Qui abbiamo (1) un agente: la mamma; (2) un atteggiamento: FA FINTA; (3) un'àncora: la banana; e (4) un contenuto: la proposizione espressa dall'enunciato 'è un telefono'. In modo analogo, per prevedere correttamente il comportamento di Sally nel test della credenza falsa, ToMM deve costruire la seguente metarappresentazione:

> Sally CREDE (di) questa biglia (che) 'è nel cestino'.

Anche in questo caso abbiamo (1) un agente: Sally; (2) un atteggiamento: CREDE; (3) un'àncora: la biglia (4) un contenuto: la proposizione espressa dall'enunciato 'è nel cestino'.

Essendo ToMM un modulo centrale, gli input che lo attivano hanno carattere *concettuale* e non percettivo. Infatti ToMM non si applica ai movimenti corporei in quanto tali, bensì ad *azioni*, che sono movimenti volti a fini specifici. Inoltre il modulo può essere attivato oltre che dalla percezione esterna e dai meccanismi propriocettivi, anche da input linguistici (e dunque paradigmaticamente concettuali), come avviene quando qualcuno ci descrive le azioni di un'altra per-

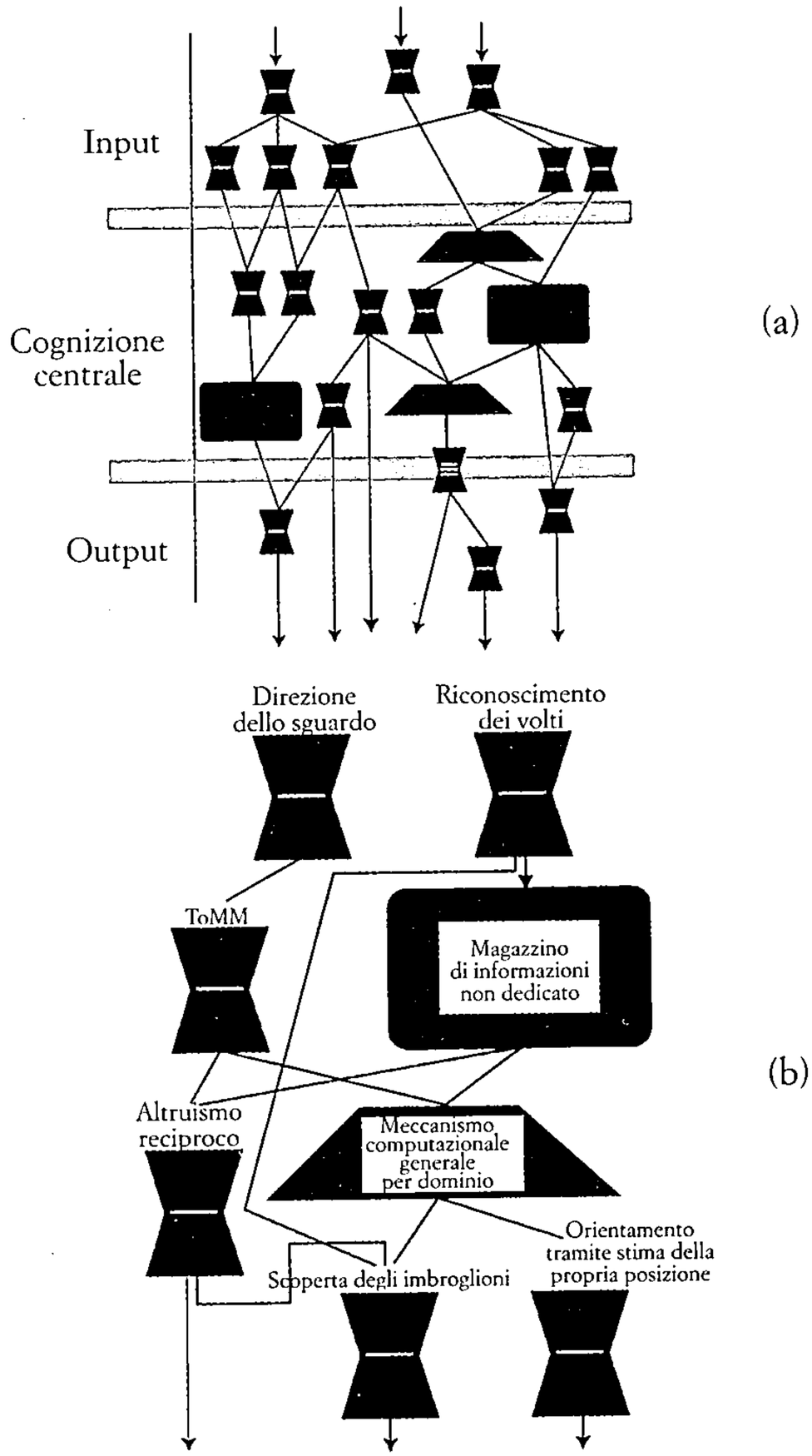


Figura 3.1. (a) Uno schizzo del tipo di architettura cognitiva di cui potrebbe far parte ToMM; (b) un ingrandimento di una parte del sistema (da Murphy e Stich 2000).

sona, o leggiamo un romanzo e cerchiamo di comprendere le azioni e i moventi dei personaggi.

Inoltre ToMM sembra possedere la proprietà cruciale dell'incapsulamento. Le forti illusioni prodotte al cinema suggeriscono un elevato grado di isolamento dei meccanismi deputati alla comprensione teleonomica e intenzionale. I soggetti dell'esperimento di Heider e Simmel (descritto nel cap. 1), pur sapendo che quelle che hanno davanti agli occhi sono figure geometriche, non possono fare a meno di mentalizzare. È del resto esperienza comune che «quando si osserva un buon attore recitare, è impossibile evitare di interpretare le sue azioni come manifestazioni di inganno, gelosia, ira ecc., pur sapendo perfettamente che l'attore non si trova in nessuno di questi stati» (Botterill e Carruthers 1999, p. 68).

ToMM in un'architettura modulare moderatamente massiva. Abbiamo visto che Leslie ritiene vi siano due moduli, ToBy e ToMM1, che costituiscono i *precursori* di ToMM2. Baron-Cohen (1995) ha sviluppato questa idea, collocando ToMM in un più ampio 'sistema di lettura della mente' (*mindreading system*). In questo sistema il processo di attribuzione di stati mentali viene scomposto in quattro meccanismi relativamente autonomi. EDD (*Eye Direction Detector*, 'rilevatore della direzione dello sguardo') rileva la presenza e la direzione di stimoli aventi forma d'occhio e interpreta come 'vedere' la relazione sussistente tra gli occhi e ciò verso cui sono diretti. ID (*Intentionality Detector*, 'rilevatore di intenzionalità') determina quando un agente agisce in vista di uno scopo. SAM (*Shared Attention Mechanism*, 'meccanismo dell'attenzione condivisa') identifica le situazioni in cui l'agente stesso e un altro soggetto stanno prestando attenzione allo stesso oggetto. Le rappresentazioni triadiche che costruisce sono l'input di ToMM.

Il sistema di lettura della mente situa ToMM all'interno di una rete di meccanismi modulari in cui l'output di un modulo funge da input (o parte dell'input) per uno o più moduli che si trovano 'più a valle'. Tale sistema può essere a sua volta situato all'interno di una più vasta rete costituita dall'in-

terconnessione di meccanismi modulari e non modulari. Ne emerge un'architettura modulare *moderatamente massiva* in cui i meccanismi modulari coesistono con meccanismi generali per dominio, con basi di dati non dedicate e strutture cognitive di tipo ancora diverso (Fig. 3.1).

Un meccanismo generale per dominio di cui ToMM necessita è il cosiddetto *elaboratore di selezione*. Secondo Leslie e Thaiss (1992), infatti, ToMM attribuisce *automaticamente* credenze dotate di un contenuto vero. Tale strategia, pur essendo ragionevole in quanto le credenze tendono a essere vere, ha un costo: nelle situazioni in cui la credenza è falsa, è necessario *inibire* la risposta automatica di ToMM. Questo è il compito dell'elaboratore di selezione. Nel test di Sally e Ann, affinché ToMM sia in grado di inferire il contenuto della credenza erronea di Sally circa la collocazione della biglia, l'elaboratore di selezione inibisce la risposta iniziale basata sulla collocazione effettiva della biglia. Questa inibizione consente al bambino di sganciare l'attenzione da una situazione reale e di riagganciarla alla situazione in quanto rappresentata da Sally.

L'ipotesi dell'elaboratore di selezione consente di spiegare perché il bambino che non ha ancora compiuto quattro anni fallisce il compito della credenza falsa ma è in grado di fornire una raffinata prestazione mentalistica nel gioco di finzione. L'incapacità di superare il test della credenza falsa è imputabile al fatto che ToMM non può ancora contare su un elaboratore di selezione *maturo*. Nel caso del gioco di finzione, invece, ToMM non necessita dell'intervento dell'elaboratore dal momento che la modificazione della realtà ('questa banana è un telefono') è decisa dal bambino stesso o da qualcuno che giocando con lui la evidenzia con il suo comportamento.

L'ipotesi dell'elaboratore di selezione trova conferma in alcuni esperimenti che hanno posto a confronto soggetti autistici e normali. Nel 1985 Baron-Cohen, Leslie e Frith hanno avanzato la fortunata ipotesi che la sindrome autistica sia, almeno in parte, riconducibile a una compromissione di ToMM. In questo lavoro tre gruppi di bambini furono sottoposti al test di Sally e Ann: bambini autistici (età media: 11 anni), bambini con sindrome di Down (età media: 10 anni) e

bambini normali di 4 anni. La domanda critica, quella che richiede di ragionare sulla rappresentazione mentale di Sally, fu superata dalla maggioranza dei bambini normali e dei bambini con sindrome di Down, ma solo da 4 dei 20 bambini autistici. I bambini autistici fornirono dunque prestazioni di gran lunga inferiori ai bambini dei gruppi di controllo, nonostante le loro prestazioni nei test di intelligenza fossero state migliori di quelle dei bambini Down.

In seguito Leslie e Thaiss (1992) hanno indagato la comprensione di credenze false e la comprensione di rappresentazioni pubbliche erronee in bambini normali e autistici. In entrambi i tipi di comprensione la rappresentazione su cui deve concentrarsi il bambino raffigura in modo scorretto uno stato di cose. Ad esempio, nel 'compito della fotografia datata' (*out of date photo task*) viene data al bambino una macchina fotografica polaroid, con cui scatta una foto a una cameretta giocattolo in cui c'è un gatto sopra una sedia. Prima che la foto si sviluppi, lo sperimentatore la nasconde dietro la schiena. Il gatto viene quindi spostato sopra il letto e al bambino viene posta la domanda test sulla posizione del gatto *nella foto*. Al bambino si chiede pertanto di sganciare l'attenzione dalla situazione reale, la situazione che il bambino percepisce e rappresenta come vera, e focalizzarsi sulla situazione rappresentata erroneamente dalla foto. I risultati ottenuti sono i seguenti. I bambini normali di tre anni falliscono sia questo compito che quello della credenza falsa; il bambino autistico è in grado di svolgere perfettamente il compito della foto datata ma fallisce quello della credenza falsa; i bambini di quattro anni sono in grado di svolgere correttamente ambedue i compiti. Queste differenti prestazioni sono spiegabili ipotizzando che il bambino di tre anni disponga di un ToMM intatto ma di un elaboratore di selezione ancora immaturo; che il bambino autistico abbia un ToMM compromesso ma un elaboratore di selezione sufficientemente maturo per svolgere il compito della foto datata; che il bambino di quattro anni possa contare sia su un ToMM intatto che su un elaboratore di selezione maturo.

Questi dati confutano la tesi di Perner secondo cui la comprensione delle rappresentazioni mentali sarebbe il frut-

to dell'estensione al dominio psicologico di una precedente comprensione del funzionamento di rappresentazioni pubbliche. Infatti, se questo fosse effettivamente il caso, il soggetto autistico dovrebbe incontrare difficoltà con entrambi i tipi di rappresentazione. E difatti Perner (1993) ipotizza che la sindrome autistica sia causata da una compromissione dei meccanismi responsabili dello spostamento rapido dell'attenzione, il che impedirebbe il costituirsi di una base di dati idonea per lo sviluppo di una teoria delle rappresentazioni *in generale*. Tuttavia ciò che emerge dal test delle foto datate è che il soggetto autistico incontra gravi difficoltà nel comprendere le rappresentazioni mentali ma non quelle pubbliche. Un dato, questo, ben spiegato dalla teoria che fa derivare l'autismo dalla compromissione di un meccanismo cognitivo specifico al dominio della psicologia.

Un'ultima osservazione sulla necessità di situare ToMM all'interno di un'architettura modulare *moderatamente* massiva. ToMM e l'elaboratore di selezione rendono possibili capacità metapsicologiche fondamentali. Tuttavia, in una fase più avanzata dell'ontogenesi, queste capacità vengono reclutate da processi cognitivi superiori per compiti più complessi, e le mentalizzazioni di più alto livello che ne scaturiscono possono senz'altro interagire in modo *non* modulare con altri processi cognitivi, manifestando una variabilità transindividuale e transculturale. Con le parole di Scholl e Leslie, la facoltà metapsicologica adulta è «una collezione di svariate capacità interconnesse, ciascuna delle quali è realizzata cognitivamente da un meccanismo o da un processo distinti» (Scholl e Leslie 1999, p. 140).

ToMM come modulo non parametrizzato. Nel cap. 2 abbiamo introdotto la nozione di modulo *diacronico* e discusso il suo caso paradigmatico: il meccanismo teorizzato da Chomsky che assume in ingresso i dati dell'ambiente linguistico e, attraverso un processo di selezione dei valori di alcuni parametri della grammatica universale, produce in uscita la competenza linguistica adulta. Segal (1996) ha ipotizzato che anche il meccanismo per l'acquisizione della facoltà metapsicologica sia un modulo diacronico, basato sulla fissazione di parametri.

Questa proposta prende le mosse da un lavoro di Perner, Baker e Hutton (1994), in cui si sostiene che il bambino, prima di entrare in possesso del concetto di CREDENZA, possiede un concetto in cui coesistono indifferenziati FINZIONE (*PREtend*) e CREDENZA (*beLIEF*): il concetto di PRELIEF, che consente di comprendere che una persona può agire *come se* qualcosa fosse in un certo modo (come se 'la banana fosse un telefono') quando in realtà non lo è. Secondo Segal avremmo qui il parametro di un modulo diacronico associato a due opzioni alternative: PRELIEF e BELIEF. Intorno ai 4 anni di età il bambino fissa il parametro selezionando l'opzione BELIEF.

Tuttavia questa descrizione dell'acquisizione della facoltà metapsicologica come processo di fissazione di un parametro in un modulo diacronico è assai poco plausibile. La selezione di parametri è necessaria soltanto nel caso in cui è impossibile fissare *a priori* lo stato finale della facoltà da acquisire. Nel caso del linguaggio, ad esempio, vi sono vari stati finali cui deve pervenire la competenza linguistica, corrispondenti alle diverse lingue naturali. Ecco perché deve esserci un parametro del soggetto nullo: fissare in modo innato questa variabile comporterebbe l'esclusione di un'intera classe di lingue naturali apprendibili. Nel caso della facoltà metapsicologica sembra esservi però un unico stato finale: è assai plausibile, infatti, che l'acquisizione di tale competenza sia in larga misura uniforme tra gli individui e le culture. (Gli studi che hanno cercato di affermare l'esistenza di differenze transculturali nella facoltà metapsicologica riguardano esclusivamente aspetti molto avanzati di questa – ad esempio, differenze concernenti le credenze religiose o fenomeni come la stregoneria, la magia e il karma.) Ma se vi è un unico stato finale della facoltà metapsicologica, diviene inutile ipotizzare un processo di selezione dei parametri, essendo sufficiente postulare la maturazione endogena di un modulo *non parametrizzato*.

Questa è la strada presa da Scholl e Leslie (1999). A loro giudizio ToMM è un modulo non parametrizzato che rende disponibile la nozione di CREDENZA già a un anno e mezzo di

età. In questa prospettiva, il successo nel gioco di finzione e il fallimento nel test della credenza falsa da parte dei bambini che non hanno ancora compiuto quattro anni sono spiegabili senza far ricorso all'ipotesi piuttosto arzigogolata della fissazione del parametro PRELIEF/BELIEF: è sufficiente ipotizzare – come si è visto sopra – uno sfasamento fra il ritmo di maturazione di ToMM e quello dell'elaboratore di selezione.

1.4. *L'asimmetria tra la prima e la terza persona*

Al termine del cap. 1 abbiamo presentato la tesi secondo cui dietro all'illusione di un accesso introspettivo diretto si cela un'attività inferenziale basata su teorie esplicative socialmente condivise. In psicologia dello sviluppo una versione di questa idea è stata riproposta da alcuni sostenitori della teoria della teoria. Ecco un esempio:

> Anche se ci sembra di percepire direttamente i nostri stati mentali, questa percezione diretta è un'illusione. In realtà, la conoscenza di noi stessi, al pari della conoscenza degli altri, è il risultato di una teoria (Gopnik e Meltzoff 1994, p. 168).

D'ora in poi denomineremo *tesi della simmetria* l'ipotesi che l'attribuzione di stati mentali a se stessi (l'autoconsapevolezza) sia basata sulla medesima teoria psicologica ingenua che rende possibile l'attribuzione di stati mentali agli altri.

Contro la tesi della simmetria sono state recentemente elaborate due ipotesi cognitive sull'introspezione che, seppure in modo assai diverso, negano che l'accesso alla propria mente sia mediato da una teoria. La prima ipotesi – la più fragile – è stata elaborata nel quadro della teoria della *simulazione mentale*. Secondo questa teoria – introdotta nel 1986 da Robert Gordon e Jane Heal, e quindi sviluppata da Alvin Goldman, Paul Harris e altri – la facoltà metapsicologica di un agente poggia non già su una teoria ma su una capacità *pratica*, consistente nel *simulare* l'eziologia psicologica del comportamento altrui. E di norma ciò avverrebbe im-

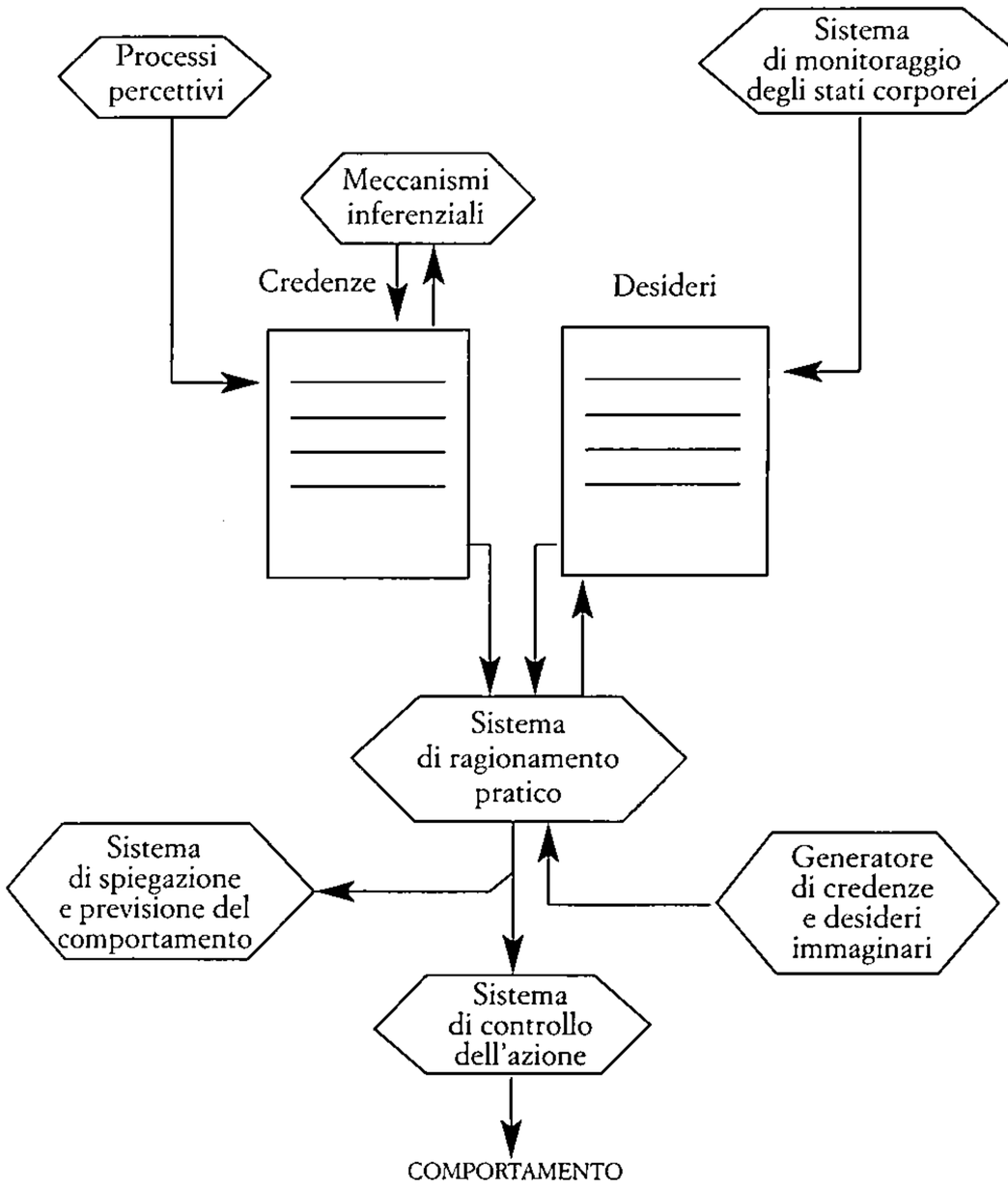


Figura 3.2. I meccanismi della simulazione mentale (da Stich e Nichols 2003).

piegando un meccanismo – il *sistema di ragionamento pratico* – reclutato per assolvere una funzione diversa da quella che solitamente svolge (ovverosia pianificare l'azione). Per comprendere questo punto, esaminiamo il diagramma della Fig. 3.2.

Il sistema di ragionamento pratico assume in ingresso credenze e desideri e prende una decisione relativa all'azione da compiere. La decisione è poi trasmessa ai *sistemi di con-*

trollo dell'azione, ossia ai meccanismi mentali responsabili del comportamento necessario per mettere in atto la decisione. Supponiamo che vi sia un meccanismo – il *sistema di spiegazione e previsione del comportamento* – in grado di mettere il sistema di ragionamento pratico 'fuori linea' (*off-line*), disimpegnandolo dai sistemi di controllo dell'azione. Supponiamo inoltre che, in questa modalità fuori linea, il sistema di spiegazione e previsione del comportamento sia in grado di alimentare il sistema di ragionamento pratico con alcune credenze e desideri ipotetici o 'immaginari' – credenze e desideri che il simulatore non ha realmente, ma che potrebbe avere l'agente il cui comportamento si sta cercando di prevedere. Se il sistema di ragionamento pratico del simulatore è simile a quello dell'agente simulato, e se le credenze e i desideri ipotetici con cui il simulatore alimenta il sistema di ragionamento pratico, si approssimano a quelli dell'agente simulato, la decisione generata dal sistema di ragionamento pratico del simulatore sarà spesso simile a quella generata dal sistema di ragionamento pratico dell'agente simulato. In quest'ottica, non vi è bisogno di una base di conoscenze rappresentate internamente; non viene impiegata alcuna teoria psicologica ingenua. Il simulatore sta impiegando una parte del proprio apparato cognitivo quale modello di una parte dell'apparato cognitivo dell'agente simulato.

Ora, mentre il teorico della teoria ritiene che gli stati mentali siano attribuiti 'in terza persona' (ad esempio, 'egli crede che *p*') da un osservatore che spiega e predice la condotta altrui in virtù di una teoria psicologica ingenua, alcuni teorici della simulazione (ad esempio Goldman 1993) ritengono che gli stati mentali siano prima di tutto attribuiti in prima persona dal soggetto a se stesso (ad esempio, 'io credo che *p*'), e quindi da lui proiettati sugli altri, 'mettendosi nei loro panni'. In tal caso, contrariamente a quanto asseriscono i teorici della teoria, la teoria psicologica ingenua dipenderebbe dall'introspezione piuttosto che costituirne la base.

Sia questa tesi simulazionista sia quella della simmetria sono state criticate da Nichols e Stich (2001). Innanzitutto i

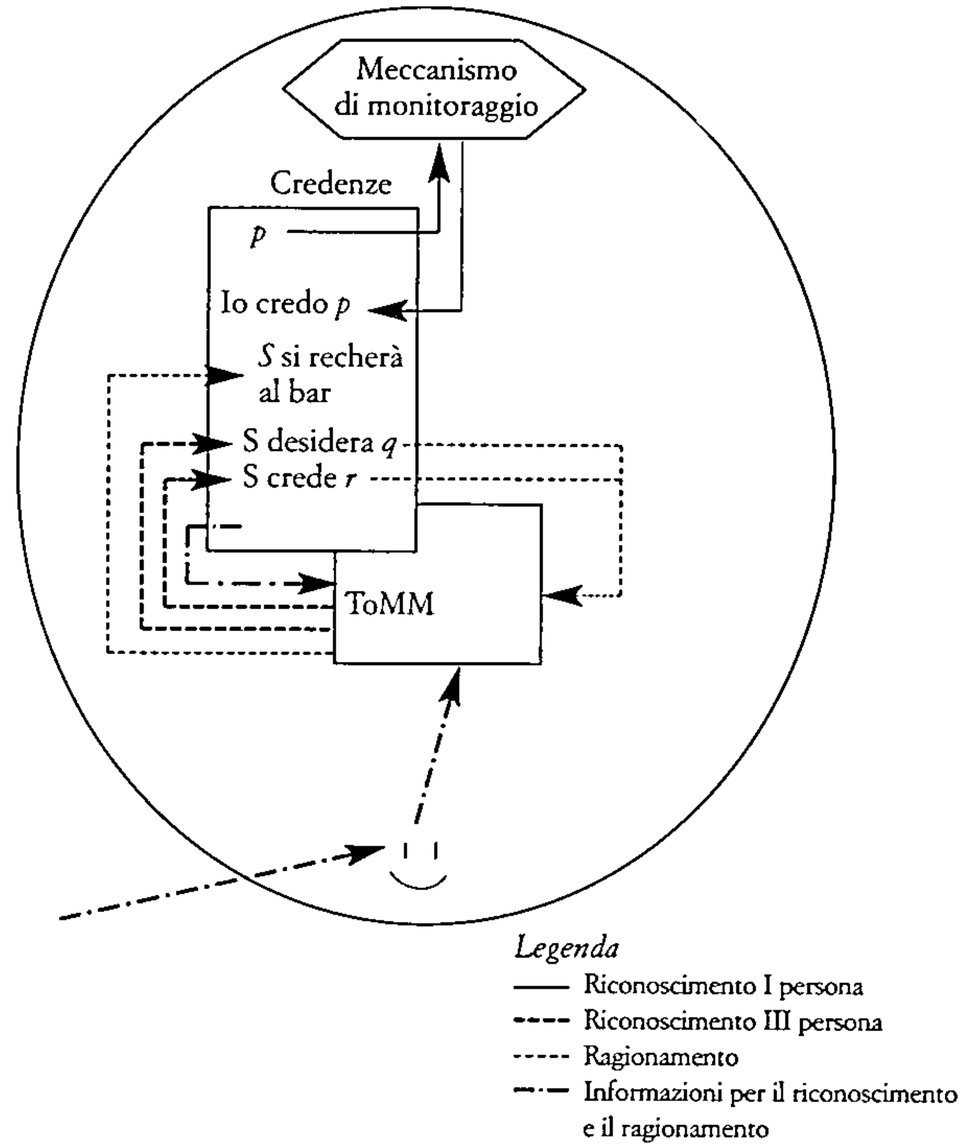


Figura 3.3. La teoria del meccanismo di monitoraggio (da Nichols e Stich 2001).

due studiosi distinguono tra l'attribuzione di uno stato mentale e le *inferenze* che su tali stati si possono compiere. Ad esempio, osservando il comportamento di una persona *P* nel contesto di una determinata situazione, posso riconoscerne gli stati mentali: *P* desidera una birra e crede che il posto più vicino per procurarsela sia il bar dietro l'angolo. Questi stati mentali possono poi essere utilizzati per fare delle inferenze: se *P* desidera una birra ed è convinto di potersela procurare recandosi nel bar dietro l'angolo, allora *P* si recherà in tale bar.

Sia il riconoscimento di uno stato mentale sia le inferenze che su di esso vengono compiute possono essere riferiti alla prima o alla terza persona. Posso riconoscere di avere uno stato mentale e quindi utilizzare questa consapevolezza nel mio ragionamento; oppure, come nell'esempio appena fatto, posso attribuire a qualcun altro uno stato mentale e avvalermene per svolgere su di esso un ragionamento. Abbiamo pertanto quattro casi possibili: riconoscimento alla prima e alla terza persona, ragionamento alla prima e alla terza persona.

Secondo Nichols e Stich la mente dispone di un *meccanismo di monitoraggio* (d'ora in poi MM) deputato all'identificazione dei propri stati mentali e del tutto indipendente dai meccanismi impiegati per identificare gli stati mentali altrui. MM assume come input uno stato mentale e genera in uscita la credenza che si è in questo stato mentale. Dunque, ad esempio, se *S* crede che *p* e si attiva MM, il meccanismo copia la rappresentazione *p* contenuta nella 'scatola' delle credenze di *S*, 'incassa' questa copia in uno schema rappresentazionale della forma 'io credo che___', e inserisce questa nuova rappresentazione nella scatola delle credenze di *S*. Si tratta, come si vede, di un meccanismo computazionale molto semplice.

MM è dunque un meccanismo introspettivo indipendente, deputato al riconoscimento dei propri stati intenzionali. Invece l'eteroattribuzione di stati mentali e il ragionamento sia alla prima che alla terza persona si baserebbero sui meccanismi della teoria della mente (Fig. 3.3).

Il meccanismo di monitoraggio si conforma abbastanza strettamente alla nozione di modulo fodoriano: è specifico a un dominio, la sua ontogenesi è precoce e prevedibile, sembra essere danneggiabile in modo selettivo[1]. Inoltre le rapide elaborazioni di MM si limitano a inserire una rappresentazione mentale in uno schema di autoattribuzione. Sotto questo aspetto, il funzionamento di MM assomiglia molto all'operare 'ottuso' dei moduli percettivi incapsulati.

[1] Nichols e Stich (2002) presentano alcune prove dell'esistenza di una doppia dissociazione tra schizofrenici e autistici: nei primi MM sarebbe compromesso ma ToMM intatto; nei secondi si osserverebbe lo schema opposto.

Forti dell'ipotesi del modulo di monitoraggio e della distinzione tra riconoscimento e ragionamento, Nichols e Stich ritengono di poter dar conto sia dei successi dell'autoattribuzione (se qualcuno mi chiedesse ora 'a che stai pensando?', non avrei difficoltà a riferire accuratamente i miei pensieri); sia degli insuccessi dell'autoattribuzione scoperti dagli psicologi sociali. Mentre nel primo caso il meccanismo soggiacente all'autoattribuzione sarebbe MM, nel secondo caso sarebbe in gioco un'attività di ragionamento alla prima persona riconducibile a ToMM.

2. *Razionalità normativa e razionalità naturale*

La discussione fin qui condotta ha fornito alcuni argomenti in favore dell'esistenza di ToMM, un modulo chomskiano-computazionale concettuale ma in una certa misura incapsulato, che ha come proprio dominio le informazioni relative agli stati mentali. Tale modulo è situato in un sistema più ampio, in cui potrebbe coesistere con il modulo di monitoraggio e i moduli ID, EDD e SAM. Questo sistema di moduli è a sua volta inserito in un'architettura modulare moderatamente massiva, popolata da meccanismi che sono sia 'dedicati' (*special purpose*) sia 'a scopo generico' (*general purpose*).

Ci volgiamo ora a un altro campo con cui si è cimentato il programma di estensione della modularità alla cognizione centrale: la psicologia del ragionamento.

Da almeno trent'anni la psicologia del ragionamento va accumulando risultati sperimentali che pongono problemi profondi alla visione classica dell'uomo come animale razionale. Infatti, ciò che questi studi sembrano dimostrare è che, in circostanze del tutto ordinarie, gli individui ragionano in modi che danno luogo a esiti contronormativi. Ne è scaturito un intenso dibattito in cui si possono distinguere varie interpretazioni degli errori inferenziali dei soggetti sperimentali.

Una interpretazione – che definiremo *pessimistica* – asse-

risce che questi soggetti giudicano e agiscono *irrazionalmente* poiché ignorano sistematicamente i principi normativi della razionalità stabiliti dalla logica deduttiva, dalla statistica matematica e dalla teoria dell'utilità attesa. A questa accusa di irrazionalità si è reagito in vari modi. In questa sede ci limiteremo a prendere in esame la reazione degli psicologi evoluzionisti. Assumendo la modularità massiva e il valore adattativo di una buona prestazione in vari compiti di ragionamento, questi psicologi sostengono che l'uomo *deve* aver evoluto moduli che risolvono compiti di ragionamento in modo normativamente corretto. Da questo punto di vista, gli errori inferenziali sarebbero imputabili al fatto che i dati forniti ai soggetti sperimentali non rappresentano l'input idoneo ad attivare i moduli in questione. Questa idea è stata definita – alludendo al celebre personaggio voltairiano secondo il quale 'tutto va per il meglio nel migliore dei mondi possibili' – l'*interpretazione panglossiana degli errori di ragionamento* (Samuels *et al.* 1999, p. 109).

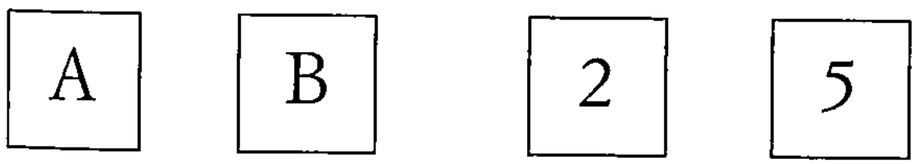
2.1. *Alcune preoccupanti prestazioni contronormative*

Un paradigma teorico molto influente in psicologia del ragionamento spiega le capacità inferenziali degli individui che non hanno mai studiato logica, attribuendo loro una *competenza logica naturale*, vale a dire un insieme di regole formali di inferenza depositate in memoria (Rips 1994). Tuttavia, qualora si prendano in esame le prestazioni medie prodotte da soggetti adulti in compiti di ragionamento su enunciati condizionali (uno dei domini deduttivi più frequentati dagli psicologi del ragionamento), si constata quanto segue. Innanzitutto i soggetti risolvono più facilmente il *modus ponens* (se *p* allora *q*; *p*; quindi *q*) che il *modus tollens* (se *p* allora *q*; non-*q*; quindi non-*p*). Inoltre i soggetti traggono conclusioni formalmente invalide: commettono, ad esempio, la *fallacia della negazione dell'antecedente* (se *p* allora *q*; non-*p*; quindi non-*q*, mentre invece nulla ne consegue) e la *fallacia dell'affermazione del conseguente* (se *p* allora *q*; *q*; quindi *p*, mentre

invece nulla ne consegue). Come spiegare questi risultati tenendo ferma l'ipotesi della logica mentale?

Una spiegazione potrebbe essere questa: gli errori inferenziali non sono dei veri errori logici, bensì conseguenze dell'azione di vari fattori *extra logici*, che non mettono in discussione la realtà della competenza logica. Tra le fonti degli apparenti errori di ragionamento potrebbero figurare i limiti della capacità umana di memoria e di attenzione, nonché le tendenze sistematiche a valutare o produrre inferenze sulla base delle credenze. In quest'ottica, gli errori inferenziali non sono errori di competenza ma di *prestazione*.

Questa soluzione si scontra però con un problema: ci sono casi in cui la grande maggioranza dei soggetti non riesce a risolvere correttamente problemi formalmente non complessi, non richiedenti eccessivi sforzi di memoria o di attenzione e non riguardanti ambiti di credenze specifiche. Il più celebre di questi casi è il *problema di selezione* (o *delle quattro carte*) escogitato da Peter Wason (1966). Su un tavolo sono appoggiate quattro carte che recano lettere (vocali o consonanti) su un lato e numeri sull'altro lato (pari o dispari). Le carte hanno sul lato visibile, rispettivamente, una vocale, una consonante, un numero pari e un numero dispari. Il soggetto deve indicare la carta o le carte che è utile girare (in modo da vederne il lato nascosto) per stabilire se è vera o falsa la seguente regola: *se una carta ha una vocale su un lato, allora ha un numero pari sull'altro lato*. Qui sotto riportiamo i quattro casi:



La regola che i soggetti devono controllare è un semplice enunciato condizionale la cui tavola di verità è riportata qui sotto:

p	*q*	Se *p* allora *q*
V	V	V
V	F	F
F	V	V
F	F	V

Quindi il controllo della verità di un condizionale richiede la ricerca di quei casi in cui l'antecedente (*p*) è vero ma il conseguente (*q*) è falso. Pertanto la soluzione corretta del compito di selezione richiede che i soggetti osservino il seguente principio normativo del ragionamento deduttivo: 'per controllare la verità di un condizionale, esamina i casi in cui l'antecedente è vero per assicurarti che il conseguente sia vero, ed esamina i casi in cui il conseguente è falso per assicurarti che l'antecedente sia falso'.

Il principio prescrive al soggetto di controllare la carta che presenta una **A** perché qualora questa rechi un numero dispari sull'altro lato, l'antecedente è vero ma il conseguente falso; e di controllare anche la carta che presenta un **5** perché qualora abbia una vocale sull'altro lato, di nuovo l'antecedente è vero ma il conseguente falso. Tuttavia dei 128 soggetti a cui Wason nel 1966 assegnò il compito di selezione, il 90% fornì una risposta scorretta, scegliendo, oltre alla carta formalmente pertinente che presenta una **A**, anche una carta formalmente irrilevante, che presenta un **2**, e non scegliendo l'altra carta formalmente rilevante, quella con un **5**.

Consideriamo altri due casi di prestazioni contronormative, questa volta in compiti di ragionamento probabilistico. Tversky e Kahneman (1973) chiesero ad alcuni soggetti di fornire una stima intuitiva del numero di parole di sette lettere contenute in quattro pagine di testo inglese, prima della forma '- - - - - n -' e poi della forma '- - - - ing'. La maggioranza dei soggetti rispose che le parole della seconda forma erano più numerose delle parole della prima forma, a dispetto del fatto che le parole di sette lettere che terminano in 'ing' sono un *sottoinsieme* dell'insieme delle parole di sette lettere che hanno 'n' in penultima posizione.

In un altro esperimento Tversky e Kahneman (1983) presentarono ai soggetti il seguente profilo di personalità: Linda ha 31 anni, non è sposata, è estroversa e brillante. Ha una laurea in filosofia. Da studentessa era molto impegnata politicamente e partecipava a manifestazioni antinucleari. La consegna sperimentale era di mettere in ordine di probabilità le seguenti caratterizzazioni socio-professionali, assegnando il rango 1 alla caratterizzazione più probabile e il rango 8 a quella meno probabile:

- Linda insegna in una scuola elementare.
- Linda lavora in una libreria e segue corsi di yoga.
- Linda è attiva nel movimento femminista (caso F).
- Linda fa l'assistente sociale.
- Linda appartiene alla Lega delle Donne Elettrici.
- Linda è impiegata di banca (caso B).
- Linda lavora per una compagnia di assicurazioni.
- Linda è impiegata di banca ed è attiva nel movimento femminista (caso B&F).

In quasi il 90% dei casi i soggetti giudicarono che il caso B fosse meno probabile del caso B&F, violando così una regola del calcolo delle probabilità (detta 'regola di congiunzione') secondo cui la probabilità del verificarsi di un evento *A* non può essere inferiore alla probabilità del verificarsi dell'evento *A e* dell'evento (distinto) *B*. Tale regola produce il seguente principio normativo del ragionamento probabilistico: 'a un evento *A* non si deve assegnare una probabilità inferiore a quella che si assegna al verificarsi dell'evento *A e* dell'evento (distinto) *B*'. La violazione di questo principio è nota come *fallacia della congiunzione.*

2.2. *Errori inferenziali: l'interpretazione pessimistica*

L'interpretazione *pessimistica* dei risultati contronormativi ottenuti somministrando questi (e molti altri) compiti di ra-

gionamento è la seguente: gli errori compiuti dai soggetti sono errori non già di prestazione ma di *competenza*. I soggetti ragionano e giudicano secondo le modalità che gli sono consentite da una competenza inferenziale in cui non compaiono le regole normativamente corrette (quelle della logica deduttiva, della statistica matematica e della teoria dell'utilità attesa).

Kahneman e Tversky (1973), ad esempio, ritengono che non siano le regole della statistica a guidare il ragionamento probabilistico ordinario, bensì alcune *euristiche*, vale a dire strategie di risoluzione di problemi economiche ed efficaci ma non sistematiche. Dicendo che un'euristica non è sistematica si vuol dire che la sua applicazione, sebbene possa condurre alla risoluzione di un problema, non garantisce il costante conseguimento di tale esito. Talvolta, infatti, la stessa euristica può dar luogo a prestazioni che si discostano da quelle ottenibili attraverso l'applicazione dei principi normativi. Pertanto gli errori sistematici (*biases*) che possono derivare dall'attivazione di una o più euristiche misurano lo scarto tra la prestazione reale e quella normativamente corretta. Ad esempio, nel problema della stima di frequenza di parole viene attivata un'euristica, detta della *disponibilità*, consistente nell'effettuare la stima affidandosi a ciò che è più facilmente recuperabile dalla memoria. La fallacia della congiunzione viene invece imputata agli effetti di un'altra euristica, detta della *rappresentatività*, per cui riteniamo più probabile ciò che appare più 'tipico' (dato il profilo di Linda, l'essere femminista è l'eventualità che domina intuitivamente tutte le altre).

La morale pessimistica ricavabile dal modello delle euristiche e *biases* è ben espressa da Slovic, Fischhoff e Lichtenstein:

> Sembra che le persone non possiedano i corretti programmi per assolvere molti compiti di giudizio importanti [...]. Non abbiamo avuto l'opportunità di evolvere un intelletto capace di gestire concettualmente l'incertezza (Slovic *et al.* 1976, p. 174).

2.3. *Errori inferenziali: l'interpretazione panglossiana*

Negli ultimi dieci anni l'interpretazione pessimistica degli errori inferenziali è stata messa in discussione da alcuni psicologi evoluzionisti.

La psicologia evoluzionistica è un programma di ricerca in scienza cognitiva che si propone di realizzare una sintesi fra innatismo, psicologia computazionale e neodarwinismo. Esaminiamo una per una le componenti di questa sintesi. Innanzitutto la psicologia evoluzionistica è una psicologia computazionale: la mente è un meccanismo di elaborazione dell'informazione, descrivibile come «un calcolatore realizzato non già da chip di silicio ma da composti organici» (Barkow *et al.* 1992, p. 7). La struttura di questa mente computazionale è *prevalentemente innata*, ossia in larga misura o integralmente determinata da fattori genetici. In tal modo la psicologia evoluzionistica fa propria la critica chomskiana della concezione empirista secondo cui la struttura innata della mente consisterebbe in poco più che un meccanismo di apprendimento 'a scopo generico'. Inoltre le principali facoltà di questa mente computazionale innata sono *adattamenti*, essendo state forgiate dalla selezione naturale al fine di risolvere problemi posti ai nostri antenati cacciatori e raccoglitori dall'ambiente del Pleistocene (il periodo compreso tra due milioni e diecimila anni fa da cui è emerso l'*Homo Sapiens*). Computazionalismo, innatismo e adattivismo si fondono infine con una versione radicale dell'ipotesi della modularità massiva. La mente umana è prevalentemente, se non addirittura integralmente, composta da moduli computazionali, calcolatori dedicati alla risoluzione dei particolari problemi posti dall'ambiente ancestrale. Per sottolinearne la componente adattiva, Samuels (2000, p. 20) qualifica questi moduli come 'darwiniani'. Molti di questi moduli computano su basi di conoscenze specifiche di dominio, sono cioè moduli chomskiano-computazionali: non a caso ToMM è spesso citato come caso esemplare di modulo darwiniano.

Per riassumere, secondo la psicologia evoluzionistica la mente umana è composta 'massivamente' da moduli darwi-

niani, vale a dire meccanismi computazionali specifici di dominio, prodotti dalla selezione naturale e innati, che possono avere accesso (talvolta esclusivo) a basi di conoscenze specifiche di dominio.

La metodologia di ricerca della psicologia evoluzionistica consiste essenzialmente nel costruire e controllare ipotesi relative ai moduli darwiniani. Vi è dunque una prima fase – chiamiamola 'analisi evoluzionistica' – in cui si formulano ipotesi concernenti i problemi di elaborazione di informazioni posti ai nostri antenati dall'ambiente ancestrale. Questi problemi sono *adattivi* in quanto la loro soluzione può aver contribuito direttamente o indirettamente al successo riproduttivo (*fitness*). Una volta definito un problema adattivo ricorrente nell'ambiente del Pleistocene, si procede a ipotizzare un modulo dedicato alla sua soluzione.

Qui va introdotta una nota di cautela. Le analisi evoluzionistiche sono utilizzabili esclusivamente come euristiche che consentono di generare ipotesi plausibili circa i moduli; di per sé esse non offrono alcuna garanzia circa la verità di tali ipotesi. Supponiamo infatti che si sia dimostrata la plausibilità del seguente condizionale: se i nostri antenati avessero evoluto un modulo in grado di risolvere un certo problema, allora si sarebbe verificato un incremento del loro successo riproduttivo. Ciò non garantisce che *di fatto* il modulo si sia evoluto, dal momento che vi sono varie ragioni per cui la selezione naturale e le altre cause dell'evoluzione possono *non* produrre un meccanismo che accresce il successo riproduttivo (Stich 1991, cap. 3).

Una volta che l'analisi evoluzionistica è riuscita a produrre un'ipotesi plausibile, lo stadio successivo è quello del controllo dell'ipotesi, la ricerca di prove dell'effettiva presenza nella mente umana del modulo postulato.

Vediamo ora come alcuni psicologi evoluzionisti hanno applicato questa metodologia di ricerca in psicologia del ragionamento. L'analisi evoluzionistica conduce a congetturare che, dato il valore adattivo di una buona prestazione in vari compiti di ragionamento, la mente umana abbia evoluto moduli che incorporano competenze inferenziali normativamente non problematiche. Tuttavia l'adattamento di tali moduli ha avuto luogo nei confronti delle tipologie di informazione disponibili nell'ambiente ancestrale. Si può pertanto prevedere che la presentazione dell'informazione secondo modalità che possono aver rivestito un ruolo importante nella nostra storia evolutiva – l'input appropriato al modulo – darà luogo a un forte miglioramento della prestazione dei soggetti nei compiti di ragionamento.

Un esempio è l'ipotesi *frequentista*. Cosmides e Tooby (1996) ritengono plausibile che nell'ambiente ancestrale le informazioni statistiche fossero disponibili in forma di frequenze (cioè rappresentazioni di casi discreti come '5 su 20') piuttosto che nella forma di probabilità di singoli eventi (p = 0,15). I nostri antenati erano immersi in un ricco flusso di frequenze osservabili e utilizzabili per migliorare la presa di decisione (ad esempio, 'le ultime 5 delle 20 volte che mi sono recato nel canyon settentrionale, la caccia ha dato buoni frutti'). Pertanto, se davvero esiste un modulo deputato al ragionamento probabilistico, il suo input dovrebbe essere costituito da informazioni relative a frequenze.

Passando dall'analisi evoluzionistica alla fase di controllo dell'ipotesi di un modulo statistico, Gigerenzer (1994) e Cosmides e Tooby (1996) hanno sostenuto che la riformulazione frequentista di alcuni esperimenti di Kahneman e Tversky determina una facilitazione della prestazione dei soggetti in virtù del fatto che i loro moduli ricevono un input adeguato. Ad esempio, la fallacia della congiunzione non viene commessa se problemi come quello di Linda sono presentati in forma frequentista: cioè chiedendo ai soggetti di stimare non già la probabilità che *Linda* 'lavori in banca' o 'lavori in banca e sia attiva nel movimento femminista', ma la *percentuale di donne* che avendo le caratteristiche di Linda, 'lavorano in banca' o 'lavorano in banca e sono attive nel movimento femminista'. In questo caso solo il 22% dei soggetti giudica che le impiegate di banca femministe sono più numerose delle impiegate di banca (Fiedler 1988).

Consideriamo un altro esempio. Cosmides e Tooby (1992) hanno ipotizzato l'esistenza di un modulo concettua-

le specificamente deputato a scoprire 'imbroglioni sociali', ossia individui che possono aver incassato i benefici derivanti da uno scambio cooperativo senza averne pagato il costo. L'ipotesi appare plausibile: gli scambi costi-benefici caratterizzati da reciprocazione differita non sarebbero adattivi, se non fosse possibile scoprire le violazioni di questi stessi scambi (Girotto 1994, p. 228). Muovendo dall'analisi evoluzionistica alla fase del controllo sperimentale dell'ipotesi, Cosmides (1989) ha riformulato il test delle quattro carte come un problema in cui si richiede il controllo di una regola che ha la forma di *contratto sociale*, vale a dire:

Se si prende il beneficio *B*, allora bisogna pagare il costo *C*.

Le carte questa volta fornivano informazioni su quattro individui. Ogni carta rappresentava un individuo. Un lato della carta informava se un individuo aveva accettato il beneficio; l'altro lato informava se l'individuo aveva pagato il costo. Il soggetto doveva indicare la carta (o le carte) da girare per stabilire se uno di questi individui aveva violato la legge.

Individuo che ha preso B	Individuo che non ha preso B	Individuo che ha pagato C	Individuo che non ha pagato C

Secondo la Cosmides queste informazioni sociali attivano il modulo deputato al controllo degli scambi cooperativi, il quale conduce alla selezione dei casi corrispondenti ai possibili imbrogli, cioè alla selezione dei casi, formalmente corretti, dell''individuo che ha preso il beneficio *B*' e dell''individuo che non ha pagato il costo *C*'.

Questi e altri risultati della psicologia del ragionamento a orientamento evoluzionistico hanno indotto alcuni psicologi evoluzionisti a sostenere che gli errori inferenziali siano errori non già di competenza, come vuole l'interpretazione pessimistica, ma di *prestazione*, riproponendo così, in una nuova chiave, l'interpretazione dei teorici della competenza logica

naturale. In questa prospettiva, l'ipotesi della modularità massiva implica l'esistenza di una moltitudine di macchine computazionali che risolvono problemi adattativi in modi 'eleganti', vale a dire secondo modalità conformi alle teorie normative della razionalità (Cosmides e Tooby 1992, p. XIV). Questa è l'interpretazione panglossiana degli errori inferenziali.

2.4. *L'interpretazione panglossiana e l'ipotesi della modularità massiva*

La psicologia evoluzionistica del ragionamento è fortemente controversa. Ad esempio, è vero che l'ipotesi del modulo scopri-imbroglioni dà conto di tutta una serie di dati sperimentali; ma è anche vero che questi stessi dati sono spiegabili in base a ipotesi alternative che, in taluni casi, sono state confermate da ulteriori evidenze. (Un'alternativa di questo tipo è la teoria degli *schemi pragmatici di ragionamento* proposta da Cheng e Holyoak 1985.)

Ma supponiamo che l'ipotesi frequentista e l'ipotesi del modulo scopri-imbroglioni siano corrette. Resta da chiedersi se l'esistenza di moduli darwiniani a fondamento del ragionamento naturale comporti l'interpretazione panglossiana degli errori inferenziali. La risposta è negativa, per almeno tre ragioni.

(1) Nel caso del problema delle parole di sette lettere, il compito assegnato ai soggetti è stimare non già la probabilità di un evento singolo ma la *frequenza* di parole. Perché mai allora la stima viene effettuata ricorrendo all'euristica della disponibilità piuttosto che al modulo deputato a ragionare sulle frequenze?

Una risposta potrebbe essere la seguente. Dato un problema, vi è un meccanismo in grado di determinare quale modulo, fra i molti presenti in una mente massivamente modulare, vada attivato per risolverlo. È questo meccanismo di *allocazione* che invia il problema della stima della frequenza delle parole nel posto sbagliato: a un modulo che impiega

l'euristica della disponibilità per ragionare in condizioni di incertezza piuttosto che al modulo deputato a ragionare sulle frequenze (Samuels 2000, pp. 26-27). Supponiamo che questa ipotesi architettonica sia corretta e che l'allocazione erronea sia il frutto di proprietà permanenti del meccanismo di allocazione: ne dobbiamo concludere che è il meccanismo stesso ad essere problematico sotto il profilo normativo, dal momento che genera sistematicamente errori inferenziali attivando un modulo inadeguato. Tuttavia il fatto che taluni errori inferenziali siano il frutto di un meccanismo normativamente problematico comporta che, a differenza da quanto sostenuto dall'interpretazione panglossiana, almeno alcuni di questi errori *non* sono errori di prestazione. Un errore di prestazione è riconducibile all'azione perturbatrice che meccanismi quali l'attenzione o la memoria di lavoro esercitano sul funzionamento normale del sistema deputato al ragionamento, inducendolo a operare in modo inconsueto. Tuttavia, nel caso del problema delle parole di sette lettere il meccanismo di allocazione funziona normalmente: viene commesso l'errore inferenziale perché il meccanismo invia regolarmente problemi da trattare in termini di frequenze al modulo dell'euristica della disponibilità. Di conseguenza gli errori di allocazione erronea non sono caratterizzabili né come errori di prestazione né come errori di competenza (Samuels *et al.* 1999, p. 111).

(2) Come spiegare la fallacia della congiunzione alla luce dell'ipotesi della modularità massiva? Sembra che la maggioranza dei soggetti non incorra più nella fallacia una volta che il problema di Linda sia stato riformulato in termini di frequenze. Tuttavia la fallacia non è definibile come un errore di allocazione perché non esiste alcun modulo a cui il meccanismo di allocazione avrebbe potuto inviare il problema nel formato non frequentista. Ma neppure lo si può considerare un errore di competenza perché esiste un modulo che, pur non possedendo gli algoritmi per risolvere il problema nel formato non frequentista, possiede quelli per risolverlo nel formato frequentista. La prestazione dei soggetti può essere

migliorata senza intervenire sulla loro competenza, ma insegnando loro a riformulare il problema, a conferirgli cioè il formato che il modulo è in grado di elaborare. Abbiamo dunque un'ulteriore tipologia di errore: l'errore di *formulazione*.

Dunque l'interpretazione panglossiana sbaglia nel ricavare dall'ipotesi della modularità massiva la conclusione che gli errori inferenziali sono esclusivamente errori di prestazione. Alla luce di quanto si è appena detto, agli errori di competenza e di prestazione si aggiungono gli errori di allocazione erronea e di formulazione.

(3) L'interpretazione panglossiana asserisce che tutti gli errori inferenziali sono errori di prestazione perché la mente umana è costituita da *macchine eleganti*, dove 'eleganti' significa 'che ragionano secondo modalità normativamente appropriate'. Consideriamo allora il modulo per la scoperta degli imbroglioni: è una macchina elegante? Trattandosi di un modulo specifico a un dominio, prima di poterlo valutare dal punto di vista normativo è necessario identificarne il dominio. Ma qui sorge un problema. Sperber ha tracciato una distinzione fra il dominio *effettivo* e il dominio *proprio* di un modulo. Il dominio *effettivo* di un modulo è costituito da «tutte le informazioni nell'ambiente dell'organismo che possono [...] soddisfare le condizioni di input del modulo» (Sperber 1996, p. 136), dove per 'condizioni di input' devono intendersi quelle condizioni che devono essere soddisfatte affinché il modulo sia in grado di elaborare determinate informazioni. Ad esempio, se un modulo richiede che un problema sia presentato in un particolare formato, ogni informazione che non possiede tale formato non è in grado di soddisfare le condizioni di input del modulo.

Il dominio *proprio* di un modulo è costituito invece da «tutte le informazioni che il modulo ha la *funzione biologica* di elaborare» (Sperber 1996, p. 136), dove per funzione biologica di un meccanismo si deve intendere una classe di effetti che esso produce e in virtù dei quali è una proprietà stabile di una specie. In taluni casi è possibile che dominio effettivo e dominio proprio coincidano; ma in molti altri casi ciò non si verifi-

ca. Ad esempio, il dominio proprio di ToMM include esclusivamente le informazioni relative agli stati mentali umani (e al comportamento da essi causato) che possono essere risultate utili ai nostri antenati ancestrali. Tuttavia il modulo elabora anche informazioni relative a molte altre cose, ad esempio le attività di animali non umani, di personaggi dei fumetti e di oggetti fisici privi di mente (si pensi alle figure geometriche dell'esperimento di Heider e Simmel). È dunque probabile che una valutazione normativa del modulo relativizzata al suo dominio proprio sia molto più favorevole di una valutazione normativa relativizzata al suo dominio effettivo. Quale dei due domini si dovrebbe utilizzare per valutare il modulo? Quale dominio è quello corretto? Le affermazioni normative sui moduli darwiniani non hanno un senso chiaro finché non siano state relativizzate a un dominio. Si può quindi avanzare più che legittimamente il sospetto che

> i teorici panglossiani che descrivono i moduli darwiniani come 'macchine eleganti' assumono tacitamente che la valutazione normativa vada riferita al dominio proprio, mentre coloro che propongono una valutazione più pessimistica della razionalità umana riferiscono tacitamente le loro valutazioni al dominio effettivo che, nel mondo contemporaneo, contiene un gran numero di problemi di elaborazione di informazioni che sono molto diversi da qualunque cosa si sono trovati ad affrontare i nostri antenati del Pleistocene (Samuels *et al.* 1999, p. 114).

La morale che si può ricavare dalle considerazioni fin qui svolte è che l'interpretazione panglossiana non descrive correttamente le implicazioni che l'ipotesi della modularità massiva ha per lo studio del ragionamento naturale. E si può concordare con Samuels, Stich e Tremoulet quando sostengono che se l'architettura della mente è davvero massivamente modulare, e se i moduli darwiniani sono davvero a fondamento del ragionamento naturale, «molte categorie e distinzioni che filosofi e scienziati cognitivi hanno utilizzato per descrivere e valutare la cognizione dovranno essere rielaborate o abbandonate» (Samuels *et al.* 1999, p. 114).

4
Psicologia e neuroscienza

L'immagine della psicologia computazionale fin qui emersa è quella di un programma di ricerca 'progressivo', malgrado la pressante esigenza di una composizione della controversia sulla modularità. Alla pessimistica restrizione fodoriana della psicologia computazionale ai processi di input e output abbiamo visto opporsi il progetto di una psicologia evoluzionistica onnicomprensiva, basata precariamente sull'affascinante ma controversa ipotesi di un'architettura modulare integralmente massiva. Fra questi due estremi vi sono una serie di ricerche su moduli come ToMM che alimentano un cauto ottimismo circa la possibilità di estendere la portata della psicologia computazionale oltre la periferia della cognizione.

Vi è però un aspetto della psicologia computazionale che in questi ultimi vent'anni si è rivelato, senza mezzi termini, 'regressivo': l'*antibiologismo*. Fino alla fine degli anni Settanta uno dei principali argomenti offerti da Fodor a favore di una versione antibiologista della psicologia computazionale era stata l'assenza di alternative, la constatazione che essa rappresentava 'l'unico gioco lecito' in scienza cognitiva. Agli inizi degli anni Ottanta, tuttavia, questa premessa cominciò a scricchiolare. Il peso crescente che la neuroscienza andava acquisendo in scienza cognitiva rendeva urgente l'esigenza di conferire plausibilità neurobiologica ai modelli computazionali. Tale esigenza condusse a un'intensa collaborazione tra psicologi e neuroscienziati da cui gradualmente prese forma

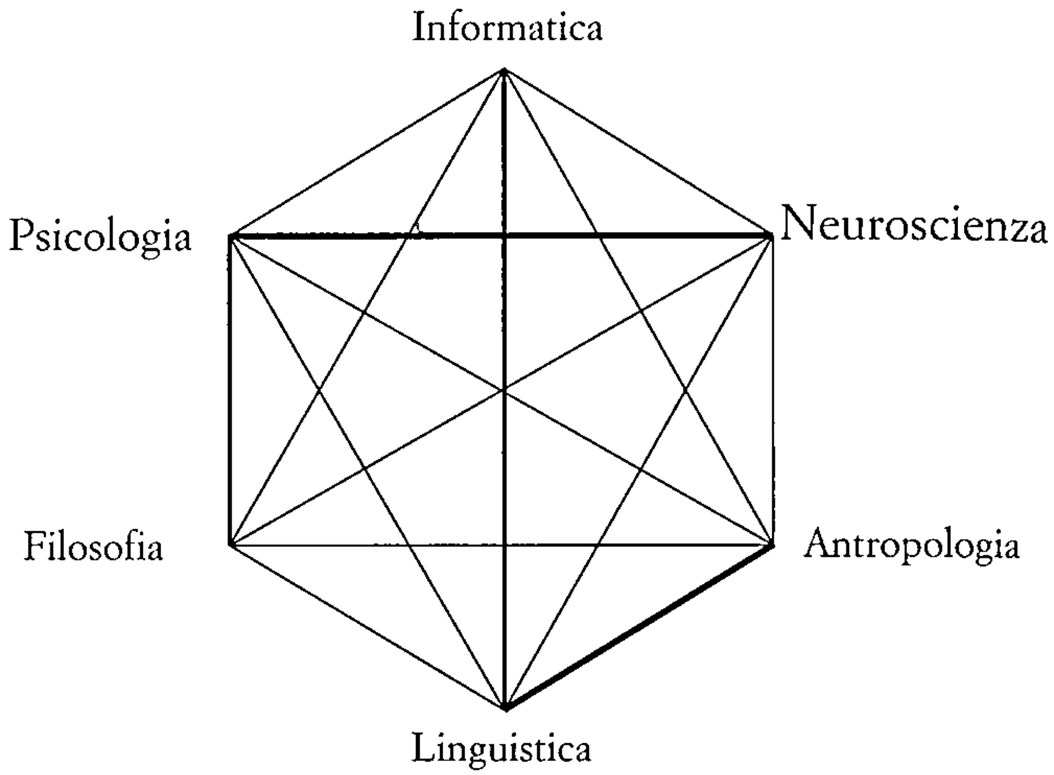


Figura 4.1. Discipline e connessioni interdisciplinari nell'attuale stadio di sviluppo della scienza cognitiva. Le dimensioni del carattere indicano l'importanza relativa della disciplina; le linee raffigurano le connessioni interdisciplinari, con lo spessore che varia in proporzione all'intensità della connessione (da Bechtel *et al.* 1998).

la *neuroscienza cognitiva*, un programma di ricerca che ha ormai spodestato l'intelligenza artificiale quale fonte primaria di ispirazione per la scienza cognitiva (Fig. 4.1).

In questo capitolo esamineremo due differenti modelli filosofici delle relazioni che psicologia e neuroscienza possono intrattenere all'interno della scienza cognitiva: il primo tenta di coniugare eliminativismo e riduzionismo psiconeuronale; il secondo propone una forma di pluralismo esplicativo.

1. *L'autonomia della psicologia*

Alla fine degli anni Cinquanta godette di un effimero momento di gloria una teoria della relazione tra il mentale e il fi-

sico nota come *teoria dell'identità di tipo*, secondo la quale ogni tipo di fenomeno psicologico è identico a qualche tipo di fenomeno neurologico. 'Tipo' si oppone qui a 'occorrenza': il dolore è inteso come tipo (una classe o proprietà) quando ci si riferisce al dolore che io e voi possiamo provare in vari momenti; è invece inteso come occorrenza (uno specifico membro della classe o un'esemplificazione della proprietà) quando il riferimento è al dolore che sto provando in questo momento. Pertanto, quando si asserisce che gli stati psicologici sono identici a stati neurologici, si pone immediatamente la questione se l'identità cui si allude riguardi i tipi o le occorrenze di tali stati. La teoria dell'identità di tipo asserisce che tutte le occorrenze di un particolare tipo di stato psicologico (ad esempio, sperimentare un certo tipo di dolore) sono identiche alle occorrenze di un tipo correlato di stato neurologico (ad esempio, l'attivazione di certe popolazioni neuronali).

Queste identità psiconeuronali sono spesso paragonate ad alcune celebri identità 'tipo-tipo' scoperte in altri ambiti della ricerca scientifica. Così facendo, si coniuga la teoria dell'identità di tipo con la nozione di *riduzione interteorica*. Nella sua formulazione classica (Nagel 1961), la riduzione di una teoria a un'altra consiste nella *deduzione* delle leggi della prima teoria (la teoria *ridotta*) dalle leggi della seconda (la teoria *riducente*). In quei casi in cui la teoria riducente e quella da ridurre impieghino vocabolari differenti per descrivere i fenomeni, si rende necessario approntare un manuale di traduzione consistente in un insieme di *leggi ponte* in grado di specificare le equivalenze fra i termini della teoria da ridurre e i termini della teoria riducente (l'esempio più sfruttato è l'equivalenza fra temperatura ed energia cinetica media delle molecole).

Nel caso della teoria dell'identità di tipo, la teoria da ridurre è la psicologia del senso comune e la teoria riducente è la neuroscienza; le leggi ponte sono identità psiconeuronali. Così intesa, la teoria dell'identità di tipo è una forma di *fisicalismo riduzionistico*. Tutti i fisicalisti convergono sulla tesi che ogni oggetto nello spazio-tempo è integralmente mate-

riale, vale a dire, è una particella elementare (protone, elettrone, quark, o quant'altro) oppure è una struttura aggregata, composta esclusivamente dalle particelle in questione. Ciò su cui sono in disaccordo è lo statuto di cui godono le proprietà di alto livello (nel nostro caso, le proprietà psicologiche) rispetto alle proprietà di livello più basso delle scienze fisiche. Il fisicalismo *riduzionistico* (nel nostro caso, il *riduzionismo psiconeuronale*) prende posizione in merito a questo problema, asserendo che le proprietà di alto livello sono riducibili – e dunque identiche – a proprietà fisiche.

Contro la teoria dell'identità di tipo Hilary Putnam (1967) ha però sollevato un'obiezione da molti giudicata decisiva. Se, come vuole il teorico dell'identità, identifichiamo il tipo psicologico 'dolore' con il tipo neurofisiologico 'attivazione delle fibre C', ne consegue che una creatura di una qualsiasi specie può provare dolore solo nel caso in cui sia dotata di fibre C e queste vengano attivate. Tuttavia un simile vincolo sulla biologia di ogni organismo in grado di provare dolore pare arbitrario. Infatti cosa, se non un ingiustificabile sciovinismo di specie, ci impedisce di attribuire lo stato di provare dolore a organismi dotati di sistemi nervosi estremamente differenti dal nostro e l'uno dall'altro? E del resto, il ricorso a esperimenti di pensiero sembra legittimare l'attribuzione di tale stato anche a organismi dotati di una biologia non terrestre (androidi la cui chimica si basi sul silicio e non sul carbonio), o addirittura a entità non biologiche (calcolatori digitali opportunamente programmati). Ogni tipo di stato psicologico pare dunque realizzabile da una molteplicità di occorrenze di stati appartenenti a tipi fisico-chimici o neurobiologici assai differenti (ad esempio, un pattern di attivazioni neuronali nel cervello di un mammifero o una certa distribuzione di correnti elettriche in una creatura composta di silicio).

Il teorico dell'identità potrebbe replicare che la proprietà fisica F con cui è identificata una proprietà psicologica P, può consistere nella *disgiunzione* di tutte le proprietà che denotano possibili realizzazioni fisiche di P. In simboli:

$$Px \leftrightarrow F_1x \vee F_2x \vee F_3x \vee \ldots \vee F_nx\,,$$

dove '$\vee$' è il simbolo della disgiunzione e '$\leftrightarrow$' si legge 'se e solo se' (bicondizionale). Ad esempio, ogni stato consistente nella credenza da parte di un soggetto S che la neve è bianca sarà identico con il fatto che S si trova nello stato cerebrale F_1 [per gli esseri umani] *o* con il fatto che S si trova nello stato cerebrale F_2 (per gli scimpanzé] *o* con il fatto che S si trova nello stato elettronico F_3 [per gli androidi T-2700] *o* ... *o* F_n [per i marziani la cui chimica si basa sul silicio]. Si deve però ricordare che il riduzionista concepisce le identità psicofisiche come le leggi ponte tra la psicologia e la fisica. E anche supponendo che una disgiunzione di proprietà possa essere considerata una proprietà genuina, è davvero implausibile supporre che possa essere una proprietà presente in una legge della fisica.

Pur revocando in dubbio la teoria dell'identità di tipo, l'argomento della realizzabilità multipla è compatibile con una forma più debole di fisicalismo nota come teoria dell'identità di *occorrenza*: ogni occorrenza di una proprietà psicologica è identica a un'occorrenza di una certa proprietà neurologica (in tal modo il fisicalismo è preservato), ma la proprietà psicologica non è una proprietà neurologica, bensì una proprietà descrivibile a un livello più astratto rispetto a quello della neuroscienza: il livello funzionalista.

In Fodor (1975) realizzabilità multipla, fisicalismo di occorrenza e funzionalismo computazionale si compongono in un unico blocco argomentativo in favore dell'*autonomia della psicologia*. Ciò che qui si intende affermando che la psicologia è autonoma dalla neuroscienza è che le categorie della prima classificano i fenomeni naturali in modo diverso dalle categorie della seconda. E per Fodor non potrebbe essere altrimenti: lo scopo dello sviluppo di scienze di livello superiore è infatti la scoperta di regolarità nomologiche diverse da quelle studiate dalle scienze di livello più basso. È l'interesse per tali regolarità a guidare il lavoro tassonomico al livello superiore; e l'uso dei termini in grado di cogliere queste regolarità prescinde dalle tassonomie dei livelli inferiori. La psicologia computazionale non costituisce un'eccezione, dal momento che il suo impianto tassonomico è im-

piegato per cogliere regolarità – le leggi intenzionali – diverse da quelle che si sforza di catturare la tassonomia della neuroscienza.

Da questa forte tesi autonomista Fodor ha ricavato una conseguenza metodologicamente negativa per il ruolo della neuroscienza all'interno della scienza cognitiva, giudicando la conoscenza dettagliata del cervello di scarsa o addirittura nessuna utilità ai fini della comprensione dei processi mentali. Una tesi, questa, in perfetta sintonia con lo stato della scienza cognitiva negli anni Settanta, quando ormai la neuroscienza – che pure aveva contribuito in modo importante alla gestazione della scienza cognitiva e ispirato i teorici dell'identità – era stata eclissata dall'intelligenza artificiale simbolica e dalla psicologia cognitiva quale fonte d'ispirazione per la comprensione filosofica della mente.

2. *Computazioni cerebriformi*

È innegabile che sino alla fine degli anni Settanta è esistito un forte iato tra i problemi e gli strumenti della neuroscienza e quelli della psicologia cognitiva e dell'intelligenza artificiale. Questo dato si rispecchia nell'idea di Fodor che le categorie della psicologia classificano i fenomeni naturali in modo diverso dalle categorie della neuroscienza in quanto le due discipline sono volte a cogliere, nella medesima realtà, regolarità nomologiche differenti. Tuttavia è possibile sostenere che questa disparità di intenti tassonomici sia nulla più che un artefatto dello stato delle ricerche sul sistema nervoso negli anni Settanta, essenzialmente incentrate su livelli di organizzazione molto bassi (ad esempio, il livello dei singoli neuroni). Con il tempo lo scarto tra neuroscienza e psicologia è andato assottigliandosi, con l'estensione della prima dai livelli molecolare e cellulare a quello dell'organizzazione del sistema visivo e dei sistemi della memoria. E se oggi la neuroscienza è in grado di lavorare a livelli di organizzazione più alti, prendendo in esame l'immagazzinamento e l'elaborazione di informazioni da parte di vaste popolazioni neuronali, diviene

legittimo sperare che essa sviluppi un apparato tassonomico congruente con gli scopi della psicologia computazionale, specialmente se questa non è la modellistica cognitiva classica bensì quella *connessionista* (P.S. Churchland 1986).

La psicologia connessionista si basa su sistemi computazionali, le *reti neurali artificiali*, che sono non già modelli di neuroni bensì «modelli *ispirati neuralmente* dei processi cognitivi» (Rumelhart e McClelland 1986, p. 180 della trad. it.). Ciò significa che, pur esistendo un innegabile scarto che separa le reti neurali artificiali da quelle biologiche, il connessionismo si caratterizza per il suo tentativo di vincolare sistematicamente la psicologia computazionale con i dati relativi all'organizzazione funzionale del sistema nervoso, vale a dire proprio con quei dati che, in accordo con l'autonomismo fodoriano, sono stati perlopiù disattesi dal cognitivismo classico.

Una rete neurale artificiale è un insieme di *unità* riunite in una fitta trama di *connessioni*. Tali unità sono dispositivi di calcolo riconducibili a tre categorie: unità di *input*, che ricevono l'informazione da elaborare; unità di *output*, in cui è possibile esaminare i risultati del processo di elaborazione dell'informazione; e unità *nascoste*, che sono collocate in uno o più strati intermedi della rete. In Fig. 4.2a è raffigurata una rete ad alimentazione anterograda, ossia una rete in cui non vi sono connessioni tra unità dello stesso livello e tutte le connessioni hanno lo stesso verso, dall'input all'output.

Le connessioni fra le unità sono paragonabili ad assoni, dendriti e sinapsi idealizzate. Ogni unità della rete simula il neurone svolgendo un'operazione molto semplice che consiste nell'attivarsi se la quantità totale di segnale che riceve supera la sua soglia di attivazione. Ciascuna unità di input è caratterizzata da uno *stato di attivazione*, un valore numerico che corrisponde alla frequenza con cui un neurone trasmette segnali attraverso le sue sinapsi. Tale valore si propaga *in parallelo* a tutte le unità nascoste cui è connessa l'unità di input. Ogni unità nascosta calcola il proprio stato di attivazione in base ai valori ricevuti dalle unità di input. Il segnale si trasmette quindi fino alle unità di output, determinandone lo stato di attivazione (la risposta della rete). La funzione della sina-

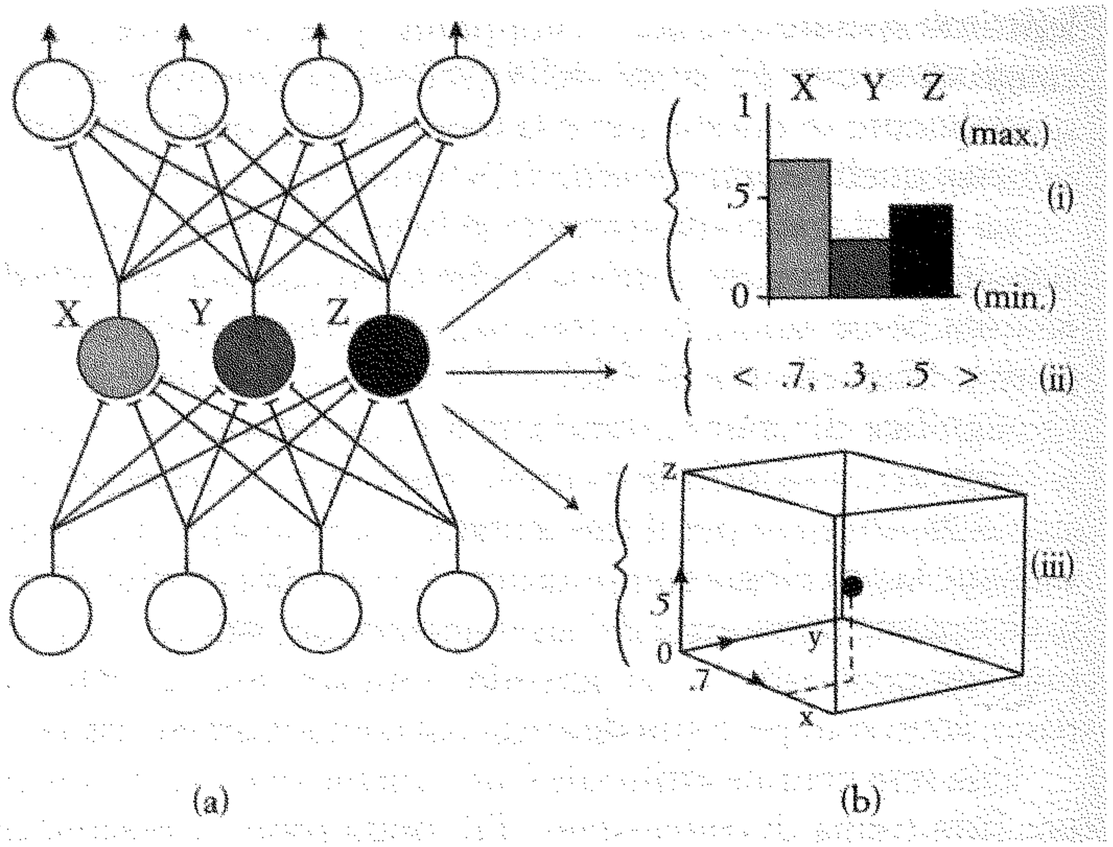


Figura 4.2. (a) Una rete neurale ad alimentazione anterograda (*feedforward*). (b) Tre modi di raffigurare i pattern dei simultanei livelli di attivazione delle unità di una rete neurale (da Churchland e Churchland 1998).

psi è simulata da un valore numerico, detto *peso*, che codifica la forza o intensità di ogni connessione. Dunque lo stato di attivazione delle unità evolve dinamicamente nel tempo, in funzione dello stato precedente e dei pesi sulle connessioni.

Per analizzare l'attività delle reti, si fa spesso ricorso ad alcune nozioni dell'algebra lineare come quelle di *vettore* e di *spazio degli stati*. La Fig. 4.2b illustra tre modi di raffigurare le configurazioni dei simultanei livelli di attivazione delle unità di una rete neurale: (i) come *istogramma* dei livelli di attivazione; (ii) come un *vettore*, <x, y, z, ... >, dove ogni elemento del vettore specifica il valore di attivazione di un'unità della rete; (iii) come un *punto* in uno *spazio degli stati*, dove ogni asse dello spazio corrisponde a un'unità della rete. Vedremo ora come (iii) consenta di costruire utili visualizzazio-

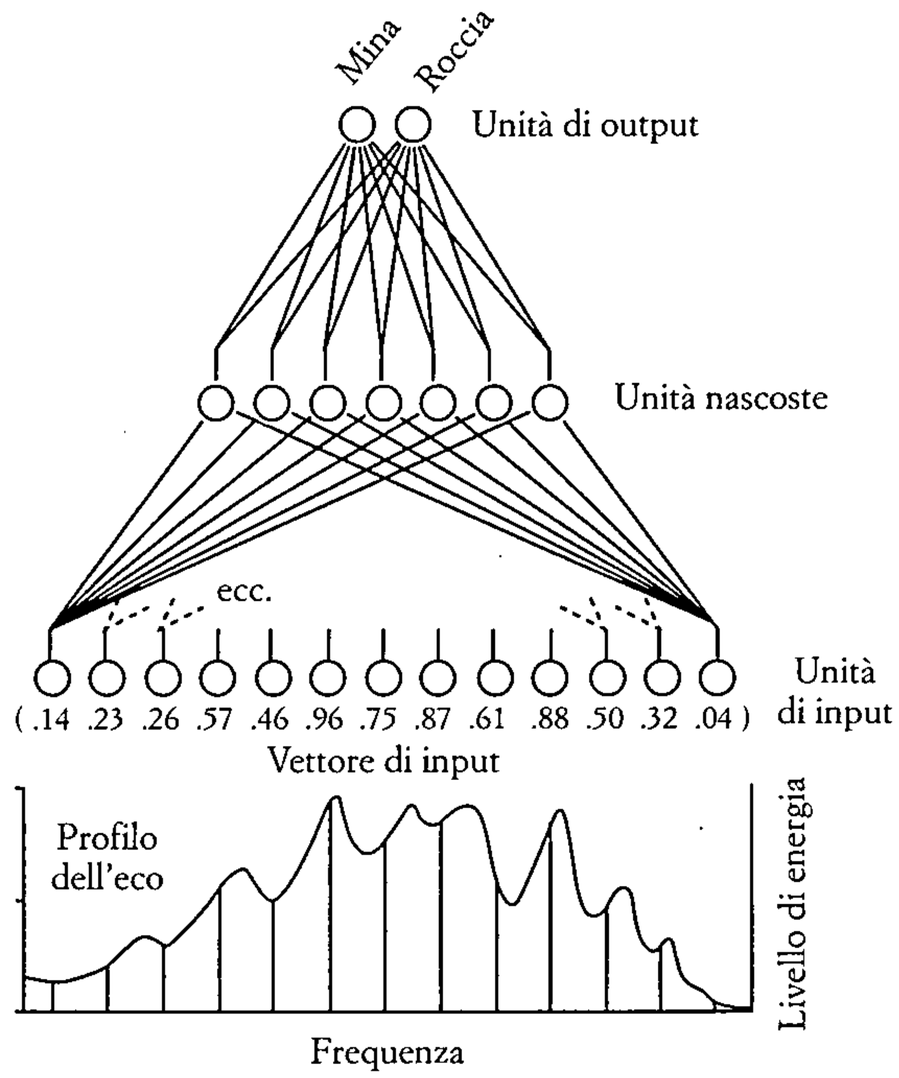


Figura 4.3. La rete che discrimina le mine e le rocce (da P.M. Churchland 1990).

ni geometriche sia del processo di apprendimento di una rete, sia della sua attività rappresentazionale.

Consideriamo la rete di Gorman e Sejnowski (1988) in grado di discriminare cilindri metallici (imitazioni di mine) e rocce sulla base dello spettro di frequenza dei loro echi rilevati da un sonar (Fig. 4.3).

In questa rete (costituita da 13 unità di input, 7 unità nascoste e 2 unità di output), il vettore di input è una rappresentazione del profilo dell'eco e l'attivazione delle due unità di output rappresenta la presenza di una mina (<1, 0>) o di una roccia (<0, 1>), rispettivamente.

Il processo di apprendimento della rete consiste nella ri-

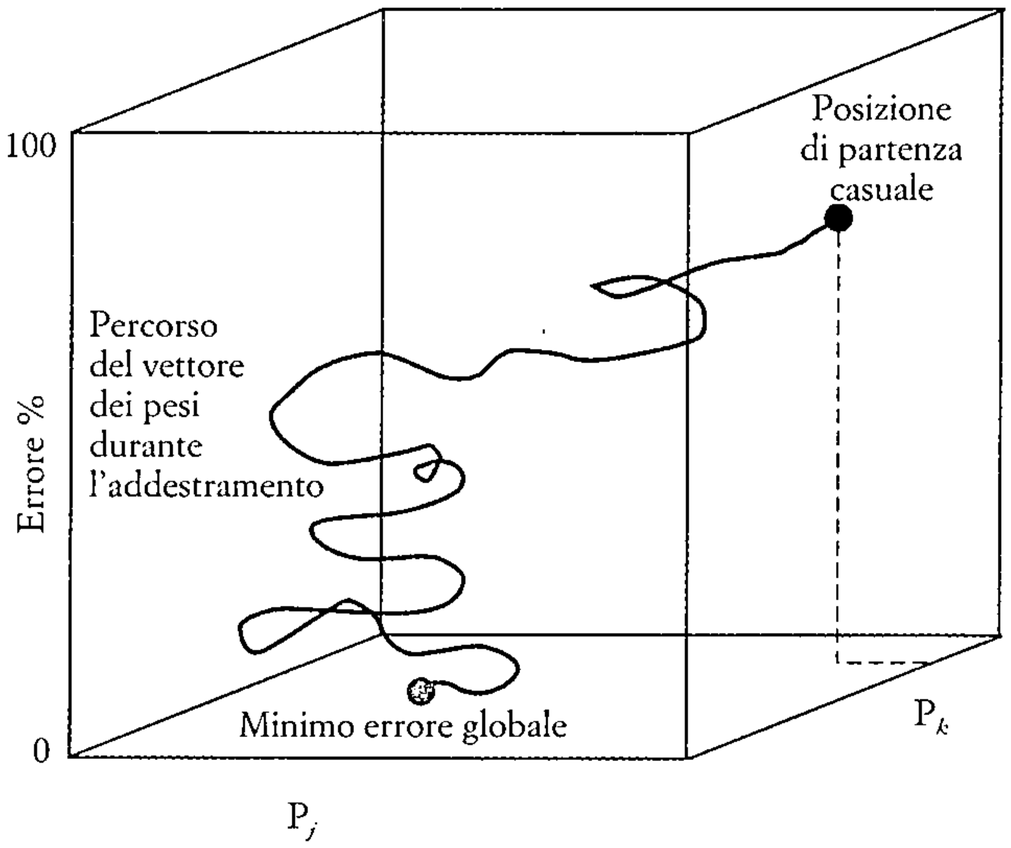


Figura 4.4. L'apprendimento di una rete rappresentato come una traiettoria di discesa nello spazio di errore dei pesi. P_j e P_k sono gli assi soltanto di 2 dei 105 pesi (da Bickle e Mandik 2002).

cerca dell'insieme dei pesi richiesto ai fini dello svolgimento del compito di riconoscimento percettivo. Questa ricerca viene condotta in base a un algoritmo di apprendimento detto di *propagazione all'indietro dell'errore*. Tale algoritmo stima l'errore della rete, calcolando lo scarto tra l'output effettivo e l'output desiderato, lo 'propaga all'indietro' dalle unità di output alle unità di input, e infine lo utilizza per modificare i pesi sulle connessioni[1]. Questo ciclo si ripete fino a quando l'entità complessiva dell'errore non si è ridotta al punto

[1] I metodi di apprendimento di una rete neurale si dividono in due classi: *supervisionato* e *non supervisionato*. Quello discusso nel testo è l'apprendimento supervisionato, cosiddetto perché esiste una specie di 'insegnante' esterno (un calcolatore) che di volta in volta 'dice' alla rete qual è la risposta desiderata.

che la rete è in grado di discriminare le mine dalle rocce. Avvalendoci della nozione di spazio degli stati, possiamo rappresentare tale processo come uno *spazio dell'errore dei pesi*, dove una dimensione rappresenta la percentuale di errore del vettore di output a ogni presentazione del vettore di input, e le altre rappresentano i 105 pesi sulle connessioni (Fig. 4.4).

Ogni punto di questo spazio degli stati rappresenta la correlazione tra l'errore complessivo del vettore di output e i valori dei pesi a ogni presentazione del vettore di input. Man mano che i pesi si modificano, l'errore globale della prestazione della rete si riduce. In tal modo, l'apprendimento è rappresentato come la correlazione fra le modificazioni dei pesi e la *discesa* lungo la dimensione dell'errore nello spazio degli stati.

Esaminiamo ora l'attività rappresentazionale in una rete neurale. Di regola i connessionisti interpretano le configurazioni dei pesi associati alle connessioni come le rappresentazioni *immagazzinate* dalla rete; i pattern dei livelli di attivazione delle unità nascoste sono visti invece come le rappresentazioni *occorrenti* (ossia le rappresentazioni che hanno luogo qui e ora). A questa distinzione se ne aggiungono altre una volta che prendiamo in esame *ciò che* viene rappresentato: concetti o proposizioni. Qui ci limiteremo a considerare le rappresentazioni connessioniste di concetti.

Si possono distinguere almeno due tipi di rappresentazioni connessioniste di concetti, a seconda che le singole unità della rete siano o meno interpretabili semanticamente. Innanzitutto vi sono reti in cui le unità su cui sono *distribuite* le rappresentazioni di livello superiore (i *concetti*), sono a loro volta rappresentazioni di livello inferiore (i *tratti distintivi* del concetto). Questo stile rappresentazionale è ben illustrato da una rete realizzata da Rumelhart, Smolensky, McClelland e Hinton (1986). La rete elabora rappresentazioni di vari tipi di stanze. Ad esempio, la Fig. 4.5 raffigura l'output della rete nel caso in cui siano state fissate inizialmente le unità divano, letto, porta e soffitto. Ciò che ne risulta può essere descritto come una camera da letto grande e originale.

In questa rete ogni unità è un simbolo, in quanto rappre-

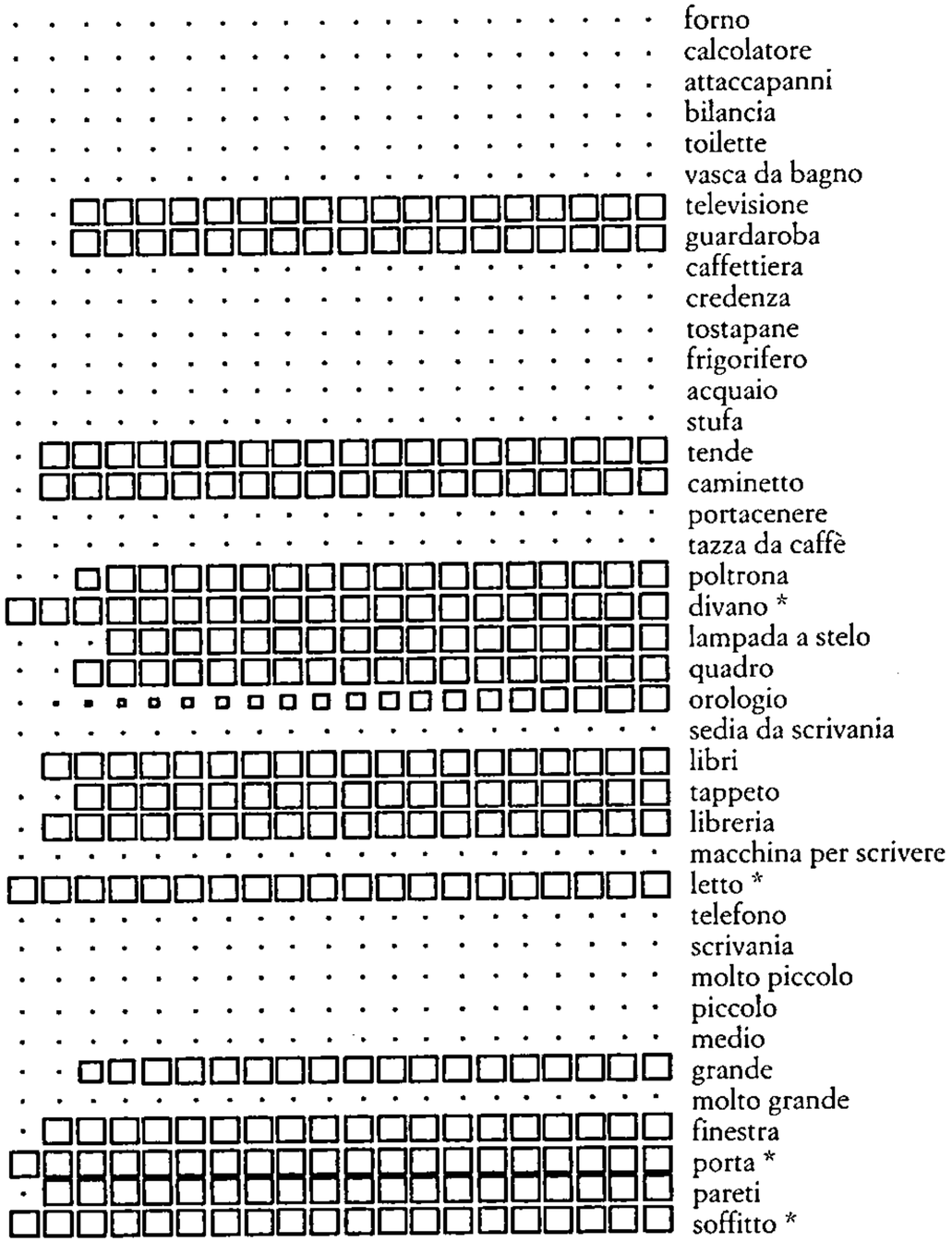


Figura 4.5. I quadratini bianchi indicano le unità attive. Gli insiemi verticali di tali quadrati (leggendo da sinistra a destra) indicano gli stati di attivazione successivi della rete. Lo stato di partenza (estrema sinistra) ha come unità attive divano, letto, porta e soffitto (da Rumelhart *et al.* 1986).

senta un tratto distintivo che può essere posseduto da una stanza. I tipi di camera (i concetti di alto livello) non sono rappresentati da singole unità della rete, ma corrispondono a

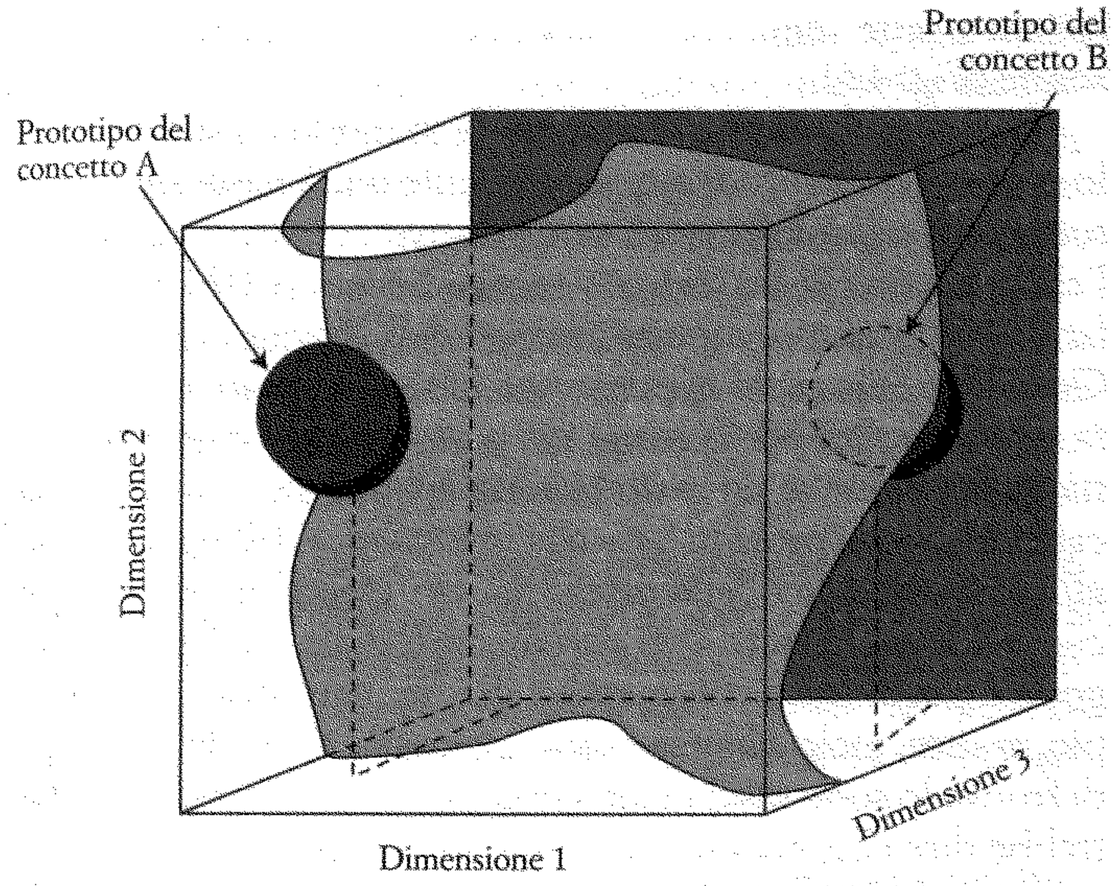


Figura 4.6. Partizione appresa dello spazio dei vettori di attivazione delle unità nascoste. Sono mostrati gli assi soltanto di tre dei sette livelli di attivazione. Concetto A = mina; concetto B = roccia (da Bickle e Mandik 2002).

pattern di attivazione in cui sono attive le unità che rappresentano le proprietà di cui gode ciascun tipo di stanza. Dunque le reti di questo tipo rappresentano i concetti attivando un insieme di unità, ognuna delle quali codifica un certo tratto distintivo che è un costituente del concetto di alto livello.

Le rappresentazioni appena discusse sono distribuite e *simboliche*. Vi sono tuttavia reti in cui le unità su cui sono distribuite le rappresentazioni dei concetti *non* sono valutabili semanticamente. In tal caso, si parla di rappresentazioni distribuite e *sub*simboliche. Troviamo questo tipo di rappresentazioni nella rete che discrimina le mine dalle rocce. Per catturare la sua attività rappresentazionale, costruiamo uno

spazio a sette dimensioni, una per ciascuna delle unità nascoste (Fig. 4.6).

La figura rivela che la rete opera in modo da suddividere lo spazio dei vettori di attivazione sulle unità nascoste in due sottospazi complementari, uno relativo ai vettori eco-mina, l'altro relativo ai vettori eco-roccia. Secondo P.M. Churchland (1989) questa interpretazione dell'attività della rete di Gorman e Sejnowski fornisce una base quantitativa e 'ispirata neuralmente' per la *teoria dei prototipi*. Secondo tale teoria, la decisione se un certo oggetto appartiene o meno a una categoria viene presa in base alla *somiglianza* dell'oggetto con gli *esemplari tipici* della categoria. Un esemplare è tipico se condivide molti tratti con altri membri della medesima categoria e pochi tratti con membri di altre categorie. Ad esempio, arance, mele e banane sono esemplari tipici del concetto di frutto; lo sono assai meno cocco, pomodoro e oliva. Si potrebbe dire che, rispetto alla categoria di appartenenza, oliva è più *periferica* o meno *centrale* di mela. Questa metrica della somiglianza è catturata dalla rete che discrimina le mine dalle rocce. Ad esempio, i vettori prossimi al centro del sottospazio relativo ai vettori-mina rappresentano esempi prototipici di echi-mina che producono un vettore di output molto vicino all'output bersaglio <1, 0>. I vettori più vicini alla superficie che suddivide lo spazio rappresentano echi-mina atipici o problematici che producono vettori di output più ambigui come <.6, .4>.

2.1. *La coevoluzione eliminativista e riduzionista*

Sotto quali aspetti il connessionismo si allontana dal computazionalismo classico? Innanzitutto il connessionismo si propone di sviluppare una versione del funzionalismo empiricamente più credibile. Al pari della psicologia computazionale classica, il connessionismo è una *psicologia* (e non una neuroscienza) perché si occupa delle forme assunte dalle rappresentazioni mentali, dei modi in cui queste possono essere manipolate e di come interagiscono fra loro allorché mediano tra

l'input sensoriale e l'output comportamentale. Il suo livello di descrizione è dunque altrettanto astratto di quello del funzionalismo computazionale classico. Ciò che il connessionismo rimprovera a quest'ultimo, tuttavia, è di non aver distinto il piano della *materia* cerebrale da quello dell'*architettura* cerebrale. Infatti, nel momento in cui il connessionista caratterizza le reti neurali come sistemi computazionali, sottoscrive *ipso facto* la tesi funzionalista secondo cui la spiegazione di un processo cognitivo astrae dal fatto che il *medium* del processo sia fatto di tessuto nervoso:

> I dettagli neuronali non sono più essenziali alle concezioni connessioniste della cognizione di quanto lo siano i dettagli relativi ai tubi a valvola o ai transistor per la concezione classica della cognizione, che permea l'intelligenza artificiale ortodossa, la psicologia fodoriana e la stessa psicologia popolare (Churchland e Churchland 1996, p. 460).

Tuttavia il connessionista sviluppa una versione del funzionalismo empiricamente più credibile quando nega che sia possibile costruire modelli computazionali prescindendo dalla struttura funzionale del cervello. Da questo punto di vista, infatti, «i pregi messi in luce dalle reti neurali [dipendono] proprio da quelle caratteristiche strutturali della computazione che sono ispirate al funzionamento del sistema nervoso centrale» (Paternoster 2002, p. 147).

Una differenza più profonda tra il computazionalismo classico e quello connessionista risiede in questo: mentre il primo segue la psicologia intenzionale nel ritenere che gli agenti cognitivi rappresentino il mondo mediante strutture simil-linguistiche ed eseguano computazioni che 'imitano' le inferenze logiche, il secondo si ispira all'organizzazione funzionale del cervello, il quale *rappresenta* il mondo mediante *vettori di attivazione*, ossia mediante pattern di livelli di attivazione su popolazioni di neuroni; ed esegue *computazioni* su queste rappresentazioni effettuando complesse *trasformazioni vettoriali* dall'una all'altra di queste popolazioni (Churchland e Churchland 1998, p. 41).

La disponibilità di una forma di psicologia computazio-

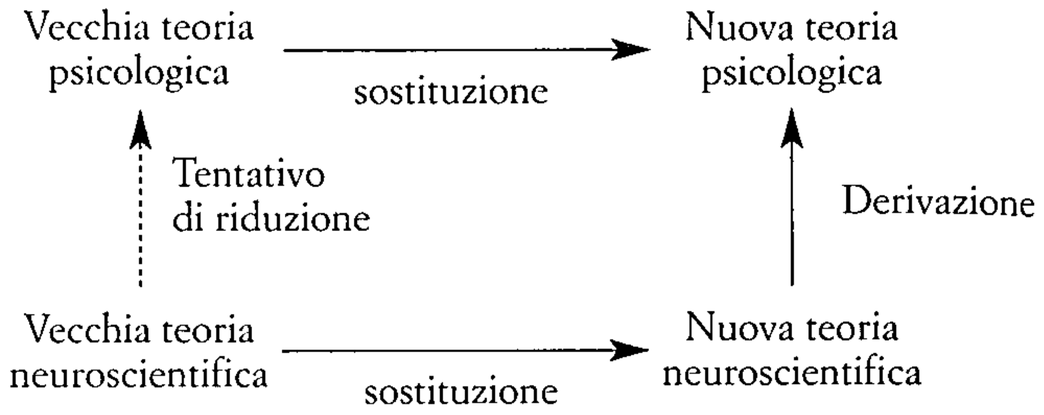


Figura 4.7. Schema di riduzione complessa: sostituzione delle vecchie teorie, psicologica e neuroscientifica, da parte delle nuove; e derivazione della nuova teoria psicologica (connessionista) dalla nuova teoria neuroscientifica (da Bechtel 1988).

nale che rompe con la 'cinematica proposizionale' e la 'dinamica logica' della psicologia del senso comune conduce i Churchland a prendere posizione sulla questione dell'autonomia della psicologia. Come sappiamo, Fodor ritiene che la psicologia computazionale sia radicalmente autonoma dalla neuroscienza in quanto i suoi stati e processi sono *irriducibili* a stati e processi neurobiologici. I Churchland accolgono la tesi dell'irriducibilità della psicologia computazionale classica, ma ne traggono una conclusione ben diversa. Essi ritengono infatti che sia necessario abbandonare una psicologia computazionale resa irriducibile dalla sua promiscuità con la psicologia intenzionale, per dedicarsi interamente a svilupparne una variante connessionista. Questa sarà allora in grado di *coevolvere* con la neuroscienza, per essere infine, in un futuro ideale, *riducibile* a una versione matura di quest'ultima. In quel momento fra i tipi funzionali della tassonomia psicologica e i tipi neurologici della tassonomia neuroscientifica si instaurerà la corrispondenza biunivoca inseguita dal riduzionismo psiconeuronale (Fig. 4.7).

Pertanto i Churchland propongono un modello a due passi – eliminativista e quindi riduzionista – della relazione tra psicologia e neuroscienza. Si badi però: il concetto di riduzione interteorica qui impiegato scaturisce da una profon-

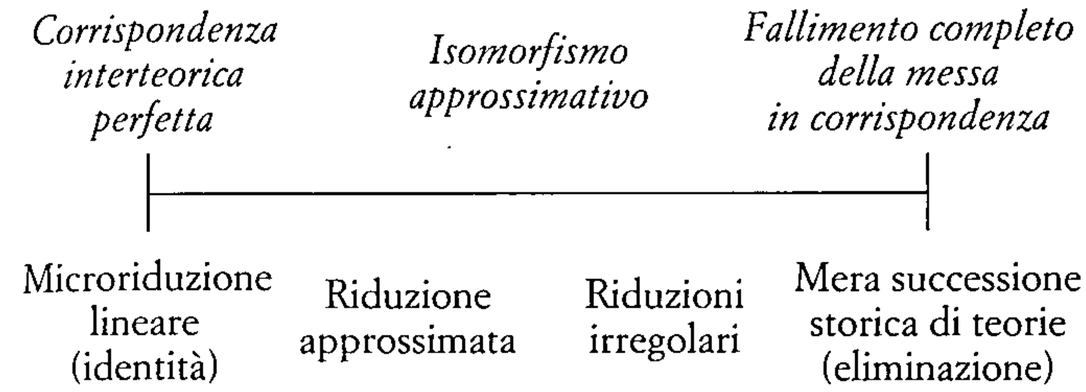


Figura 4.8. Lo spettro della commensurabilità interteorica (da Bickle 1998).

da revisione del modello classico elaborato da Nagel. Quest'ultimo è infatti incapace di dar conto di quei casi storici in cui una teoria viene sostituita o eliminata da un'altra. In tali casi non è possibile costruire leggi ponte fra la teoria riducente T_1 e la teoria da ridurre T_2 perché quest'ultima è in larga misura falsa. In termini ontologici, possiamo dire che in T_2 non vi è nulla da identificare perché T_1 ha rivelato che le entità postulate da T_2 non esistono. Di conseguenza, per fare spazio all'eliminazione i Churchland sostengono che, allorché si verifica un mutamento teorico, la 'messa in corrispondenza' fra T_1 e T_2 non viene effettuata in virtù di leggi ponte. Si costruisce invece un'*immagine equipotente* di T_2 in T_1. Si tratta di un'*immagine* (denominiamola 'T_2*') in quanto T_1 modifica e corregge T_2, preservandone soltanto alcune proprietà essenziali. Ad esempio, ciò che può essere dedotto dalla genetica molecolare non è ciò che è costitutivo della genetica mendeliana, anche se la prima incorpora proprietà della seconda quali la dominanza e la ricombinazione. Tuttavia T_2* è *equipotente* perché i principi di T_1 devono essere dotati di *tutto* il potere esplicativo e predittivo dei principi di T_2. Pertanto è T_2* (e non T_2) che è deducibile da T_1. Ciò significa che in questo caso, diversamente da quanto avveniva nel modello nageliano, la deduzione è una relazione non già *inter*teorica ma *intra*teorica. Ciò consente di dar conto non solo dei casi di riduzione in cui sono preservati i referenti dei termini di T_2, ma anche dei casi di eliminazione in cui tali refe-

renti sono stati espunti. Le corrispondenze interteoriche che in tal modo si ottengono possono essere ordinate lungo uno *spettro* che da casi di riduzione *lineare* (i casi nageliani), passando attraverso riduzioni *approssimate* e riduzioni sempre più *irregolari* (*bumpy*), conduce infine a una mera successione storica di teorie (Fig. 4.8).

Ogni qualvolta la creazione di T_2* comporta correzioni relativamente modeste di T_2, si avrà una riduzione *lineare*. Ma in quei casi in cui T_2 e T_2* sono fortemente dissimili perché, per ottenere T_2*, si è resa necessaria una massiccia revisione, diremo che T_2 è stata *eliminata*. La relazione fra l'ottica fisica e la teoria elettromagnetica di Maxwell è un esempio di riduzione lineare che ha come conseguenza ontologica la conservazione delle entità postulate dall'ottica fisica (*identità*: la luce è una radiazione elettromagnetica con una lunghezza d'onda tra i 350 e i 750 ηm). La relazione tra la chimica del flogisto a quella dell'ossidazione esemplifica invece una riduzione irregolare che ha come conseguenza ontologica la completa sostituzione della teoria del flogisto (*eliminazione*: non esiste un'entità come il flogisto).

Torniamo ora allo schema coevolutivo raffigurato in Fig. 4.7. Secondo i Churchland la prima fase prevista dallo schema è quella che la psicologia cognitiva sta vivendo a partire dai primi anni Ottanta. La constatazione che le categorie psicologico-popolari (e le categorie della psicologia cognitiva ispirate a queste ultime) non si riducono a quelle della neuroscienza, sta spingendo molti psicologi alla loro eliminazione in favore delle categorie della psicologia connessionista. Da questo punto di vista, la differenza interteorica fra, da un lato, i vettori di attivazione e le trasformazioni vettoriali connessioniste e, dall'altro lato, le rappresentazioni simil-enunciative e le inferenze simil-logiche della psicologia computazionale classica, è altrettanto grande di quella tra il concetto di flogisto e quello di ossidazione. E come si è appena visto, il flogisto è un esempio canonico di un'entità eliminata dalla nostra ontologia scientifica. La seconda fase dello schema in Fig. 4.7 è la previsione che una psicologia connessionista emancipata dalle categorie psicologico-popolari coevolverà con la neuroscienza fino al

momento in cui la prima sarà riducibile in modo più o meno *lineare* a una versione matura della seconda.

2.2. *Difficoltà dell'eliminativismo*

Vi sono almeno due obiezioni che possono essere rivolte all'eliminativismo dei Churchland. La prima è che il loro bersaglio polemico non è tanto la psicologia intenzionale quanto piuttosto la caratterizzazione che di questa ha dato Fodor. E alcuni studiosi (ad esempio Loar 1983; Stalnaker 1984) hanno contestato l'idea fodoriana secondo cui la psicologia del senso comune assegna agli atteggiamenti proposizionali una struttura sintattica simile a quella degli enunciati delle lingue naturali.

La secondo obiezione è più devastante. Esaminiamo la struttura dell'argomento eliminativista. Questo prende le mosse da una premessa secondo la quale la psicologia del senso comune è una teoria e gli stati intenzionali i suoi termini teorici (la teoria della teoria). Dopodiché aggiunge la premessa secondo cui la psicologia del senso comune è una teoria irrimediabilmente difettosa poiché molte delle sue tesi – e dei presupposti di tali tesi – sono false. Infine giunge a due conclusioni: la prima asserisce che le entità postulate dalla psicologia del senso comune non compariranno nell'ontologia di una scienza cognitiva matura; la seconda – più forte – asserisce che tali entità non esistono.

Ora, a prescindere del tutto dalla plausibilità delle premesse, è legittimo domandarsi: che cosa giustifica il passaggio dalle premesse alla seconda conclusione dell'argomento eliminativista? Perché dalla constatazione della falsità della teoria psicologica popolare si dovrebbe inferire l'inesistenza delle entità da essa postulate? È evidente che qui si richiede l'aggiunta di una premessa che incorpori qualche dottrina concernente la verità e il riferimento.

Una premessa incorporabile nell'argomento eliminativista per giustificarne la conclusione ontologica è stata individuata da Lycan (1988): l'eliminativismo presuppone una

qualche versione della teoria *descrittivista* del riferimento dei termini teorici. Stando a tale teoria, a ogni termine teorico è associato un insieme di descrizioni che ne determina il riferimento. Se, per ogni termine teorico, la totalità o la maggior parte delle descrizioni a esso associate sono soddisfatte da un'entità, il termine si riferisce a quell'entità; altrimenti il termine è privo di referente. Se la totalità o la maggior parte dei termini teorici di una teoria non possiede alcun referente, siamo costretti a concludere che le entità postulate dalla teoria non esistono.

A partire dai primi anni Settanta, tuttavia, un gran numero di critiche si sono andate accumulando a carico delle teorie descrittiviste. Un'alternativa che oggi gode di grande popolarità è costituita dalle teorie *causali* del riferimento, di cui sopra abbiamo incontrato la variante fodoriana. E tali teorie sono state utilizzate da più di uno studioso (vedi ad esempio Jacoby 1985) per sostenere che è ancora possibile fissare il riferimento dei termini mentali anche se non esiste alcunché in grado di soddisfare le proprietà attribuite loro dalla psicologia del senso comune. Pertanto, se una teoria descrittivista del riferimento è il contenuto della premessa addizionale (come accade ad esempio in P.M. Churchland 1979, pp. 56-58), l'eliminativista non può certo darla per scontata.

3. *La dialettica della scomposizione e della localizzazione*

Oltre alle difficoltà dell'eliminativismo, vi è un'altra ragione per respingere il modello delle relazioni tra psicologia e neuroscienza proposto dai Churchland: il suo carattere *unidirezionale*. Dal momento che il modello attribuisce una priorità al livello neuroscientifico, ogni volta che non è possibile istituire una chiara corrispondenza fra un tipo psicologico e un tipo neurologico, il fallimento è immediatamente imputabile alla psicologia. Tuttavia alcuni casi paradigmatici di coevoluzione della psicologia e della neuroscienza manifestano un carattere *bi*direzionale: vale a dire in talune circostanze anche considerazioni funzionali a carattere psicologico esercitano

un'influenza decisiva sulla definizione dei tipi neurologici. Per dar conto di questo fatto è necessario adottare una concezione della dinamica coevolutiva alternativa alla sintesi di riduzionismo ed eliminativismo proposta dai Churchland: una coevoluzione situata nella prospettiva del *pluralismo esplicativo*, la tesi secondo cui «realizzare simultaneamente ricerche a vari livelli di analisi della scienza tende a promuovere il progresso a ognuno di questi stessi livelli» (McCauley e Bechtel 2001, p. 738).

L'instaurarsi di connessioni fra differenti discipline scientifiche, situate a differenti livelli di analisi, offre ai cultori di tali discipline una gran mole di risorse (teoriche, pratiche e probatorie) non altrimenti disponibile. L'opportunità di usufruire di tali risorse costituisce un forte incentivo per l'esplorazione delle relazioni fra le scienze. Questo modo di vedere le cose equivale dunque a una valorizzazione pragmatica dell'incremento delle risorse esplicative. In tale quadro la *bidirezionalità* della relazione tra psicologia e neuroscienza diviene una virtù da perseguire.

È importante notare che il pluralismo esplicativo non esclude il riduzionismo psiconeuronale a condizione che questo venga inteso come una sorta di 'astuzia della ragione'. Infatti, perseguendo gli scopi dell'economia ontologica e dell'unificazione della scienza, il riduzionismo è «il più potente motore della scoperta nell'indagine transcientifica» (McCauley e Bechtel 2001, p. 774). Ad esempio, lo sforzo di stabilire equivalenze fra i termini psicologici e quelli neurobiologici (ossia di scoprire identità psiconeuronali) promuove la coevoluzione pluralista di psicologia e neuroscienza.

Un modello pluralista delle relazioni fra discipline scientifiche è quello di Darden e Maull (1987), in cui lo strumento di integrazione interdisciplinare è costituito da teorie 'ponte' (*interfield*), le quali istituiscono connessioni tra fenomeni che sono stati studiati in due o più campi di indagine, senza che per questo uno di questi campi sia più fondamentale degli altri. Tali connessioni costituiscono un'*euristica della scoperta* in quanto ciò che è noto in relazione a un certo insieme di fenomeni può promuovere ipotesi relative a un altro insie-

me di fenomeni. Un caso storico discusso dai due autori riguarda la genesi di una teoria ponte tra la citologia e la genetica. La teoria cromosomica dell'eredità mise in relazione i fattori mendeliani (che erano definiti *funzionalmente* sulla base dei tratti di cui erano responsabili) con i cromosomi, inducendo Boveri e Sutton a ipotizzare che i fattori mendeliani fossero fisicamente localizzati sui cromosomi. L'istituzione di questo nesso ha costituito il fondamento del programma della genetica classica.

Questo esempio illustra un tipo assai comune di teoria ponte, che istituisce un nesso tra entità la cui *funzione* è l'oggetto di un campo di indagine ed entità la cui *struttura* è l'oggetto di un altro campo. Attualmente teorie ponte di questo genere vengono costruite in neuroscienza cognitiva; e lo strumento con cui questa istituisce nessi tra strutture cerebrali e funzioni psicologiche è la spiegazione meccanicista.

3.1. *Due forme di scomposizione funzionale*

Vi sono almeno due modi in cui una funzione cognitiva può venir scomposta e localizzata: la scomposizione può *precedere* la localizzazione; oppure le due possono *informarsi reciprocamente*. Nel primo caso si adotta un'approccio dall'alto in basso alla scomposizione funzionale. Consideriamo, ad esempio, le tecniche di *cronometria mentale* comunemente impiegate dagli psicologi cognitivisti. Il metodo *sottrattivo* utilizza i tempi di reazione per scomporre le attività cognitive nelle loro operazioni costituenti. Un soggetto sperimentale deve eseguire due compiti, giudicati differenti in quanto – *alla luce di una scomposizione effettuata a priori* – si ritiene che un compito (ad esempio, la discriminazione tra due segnali) aggiunga un'operazione alle operazioni necessarie per espletare l'altro compito (che può essere, ad esempio, la semplice reazione a un segnale). Pertanto la differenza fra i tempi di esecuzione dei due compiti è una misura del tempo necessario per eseguire il compito addizionale. Si noti: tutto quello che abbiamo qui è un'analisi *a priori* dei compiti e una fonte

probatoria costituita esclusivamente da dati comportamentali. Lo psicologo, da buon autonomista, potrebbe giudicare del tutto superfluo collegare una simile scomposizione funzionale ai sottostanti processi cerebrali.

Anche il funzionalismo omuncolare imprime questa curvatura dall'alto in basso allo sviluppo delle spiegazioni meccaniciste. A prima vista, questa posizione sembra costituire un passo avanti rispetto a tutte quelle forme di funzionalismo che rendono la psicologia radicalmente autonoma dalla neuroscienza. Lycan (1994) lo sostiene quando afferma che la scomposizione di una funzione psicologica deve essere spinta fino al punto in cui si constata che le componenti sono *strutture neuroanatomiche*. Tuttavia, a ben guardare, la situazione non è granché mutata: l'analisi funzionale può procedere dall'alto in basso fino a quando non incontra delle strutture neuroanatomiche che sono giudicate preesistenti all'analisi stessa, vale a dire identificate indipendentemente da considerazioni funzionali. Struttura e funzione non interagiscono.

Il superamento della versione antibiologista della tesi dell'autonomia della psicologia richiede l'istituzione di un nesso più forte tra scomposizione e localizzazione. Tale nesso è presente in molte ricerche della neuroscienza cognitiva, che fanno interagire scomposizione e localizzazione in un processo di *equilibrio riflessivo*. In tal caso il modo in cui i neuroscienziati isolano le strutture neuronali (il modo stesso in cui vengono delimitate le aree cerebrali) viene a essere strettamente legato alla scomposizione delle capacità cognitive: l'assegnazione di una funzione cognitiva a una struttura neuronale è giudicata uno dei principali criteri della sua individuazione. Dall'altro lato la cartografia cerebrale diviene una guida euristica nello sviluppo di modelli di elaborazione dell'informazione, mettendo sovente capo a revisioni delle scomposizioni funzionali proposte in sede psicologica.

Gli studi di visualizzazione dell'attività cerebrale mediante tomografia a emissione di positroni (PET) sono un ottimo esempio di identificazione di strutture cerebrali in base a criteri di rilevanza funzionale, in cui cioè la relazione interdisci-

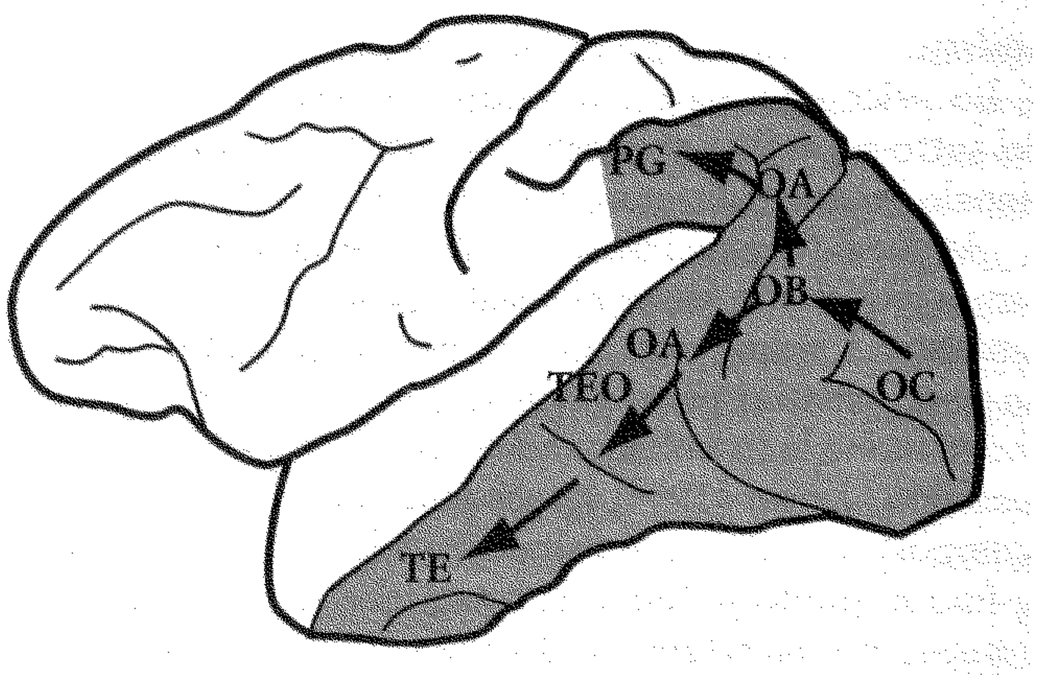


Figura 4.9. Visione laterale dell'emisfero sinistro di una scimmia *rhesus*. L'area in grigio definisce il tessuto della corteccia visiva nei lobi occipitale, temporale e parietale. Le frecce schematizzano le due vie visive corticali, ciascuna delle quali ha origine nella corteccia visiva primaria (area OC), divergendo all'interno della corteccia prestriata (aree OB e OA) per poi indirizzarsi o ventralmente nella corteccia temporale inferiore (aree TEO e TE) oppure dorsalmente nella corteccia parietale inferiore (area PG). Entrambe le vie visive sono essenziali ai fini della funzione visiva superiore: la via ventrale per la visione degli oggetti, la via dorsale per la visione spaziale (da Bechtel *et al.* 1998).

plinare va dalla psicologia alla neuroscienza. L'interpretazione funzionale delle immagini della PET è infatti ottenuta applicando (spazialmente anziché temporalmente) il metodo sottrattivo di Donders: si sottraggono le misurazioni del flusso sanguigno effettuate durante l'esecuzione di un dato compito (ad esempio, pronunciare un sostantivo visto o udito) da quelle effettuate durante l'espletamento di un altro compito che – *stando a una certa scomposizione funzionale realizzata in sede psicologica* – include tutti i passi richiesti dal primo compito più quelli peculiari al secondo (ad esempio, generare il verbo appropriato al sostantivo presentato) (Petersen *et al.* 1989).

Prendiamo ora un caso in cui le informazioni ricavate dagli studi sulle strutture neuronali determinano una revisione delle scomposizioni funzionali effettuate in sede psicologica – in cui, con buona pace di Fodor, la relazione interdisciplinare va dalla neuroscienza alla psicologia. Sulla base di dati ottenuti lesionando la corteccia del *macacus rhesus* (dunque con una tecnica specificamente neuroscientifica), Mishkin, Ungerleider e Macko (1983) hanno distinto due vie principali dell'elaborazione dell'informazione visiva. La prima è una via dorsale che dalla corteccia primaria giunge alla corteccia parietale posteriore, formando un sistema deputato alla percezione spaziale (sistema del *where*, del 'dove'). La seconda è una via ventrale che sempre dalla corteccia primaria giunge alla corteccia temporale inferiore, formando un sistema preposto alla percezione degli oggetti (sistema del *what*, del 'che cosa'). Sarebbe stato difficile ottenere questo risultato se ci si fosse limitati a un'analisi dall'alto in basso della funzione visiva (Fig. 4.9).

In seguito, la distinzione fra la via del 'che cosa' e quella del 'dove' ha svolto un ruolo euristico nella modellistica neurocomputazionale. Ad esempio, nel tentativo di comprendere perché il sistema visivo elabori separatamente informazioni relative al 'dove' e al 'che cosa', Rueckl, Cave e Kosslyn (1989) hanno realizzato una simulazione connessionista in cui la rete veniva addestrata a specificare su due insiemi distinti di unità di output, l'identità o la collocazione di immagini proiettate su una retina artificiale. Questi ricercatori hanno dimostrato che la rete apprende il compito più rapidamente se si impiegano due insiemi distinti di unità nascoste, un insieme connesso alle unità di output relative al 'dove' e l'altro collegato alle unità di output relative al 'che cosa'. In altre parole, l'apprendimento risulta più rapido quando le due parti del compito sono distribuite tra due sottoreti che hanno in comune soltanto le unità di input.

Jacobs, Jordan e Barto (1991) hanno proseguito questo lavoro al fine di mostrare che una rete può autonomamente 'decidere' di scomporre un compito, assegnando i sottocompiti a varie sottoreti. I ricercatori hanno utilizzato un sistema

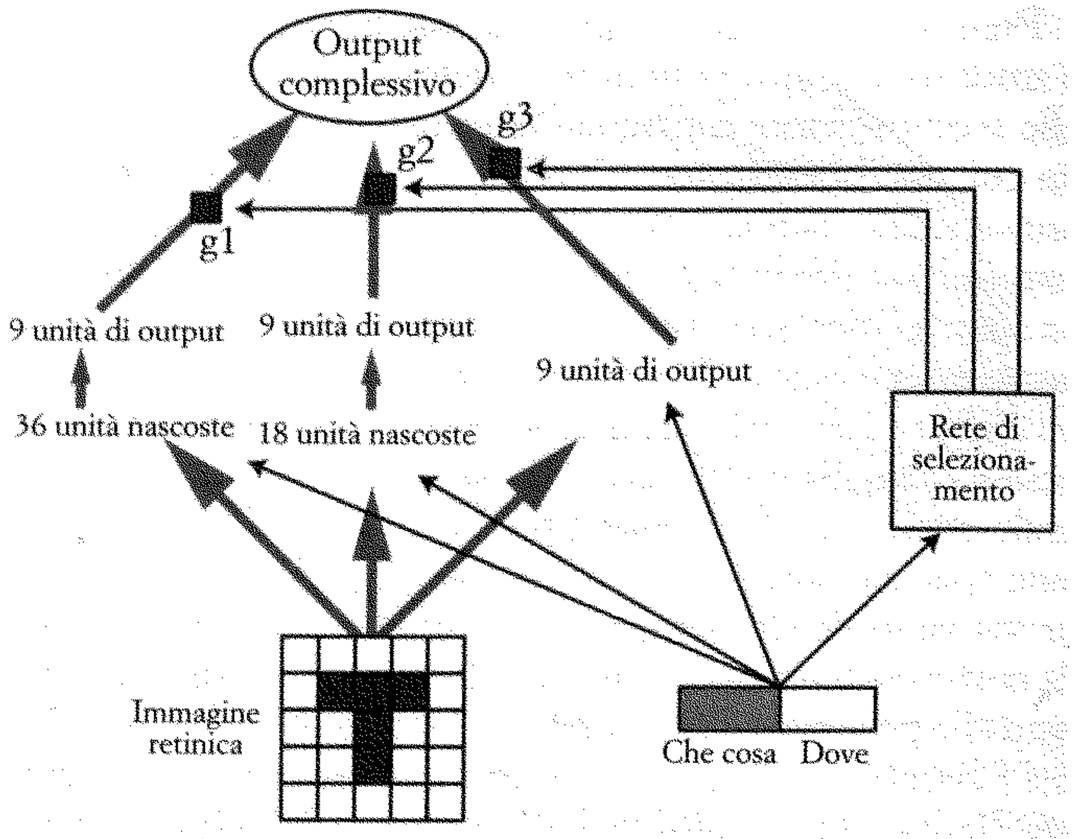


Figura 4.10. Simulazione connessionista dei compiti 'che cosa' e 'dove'. Le frecce più grandi indicano un insieme di connessioni tra le unità dei due strati; le frecce più piccole indicano un'unica connessione (da Jacobs *et al.* 1991).

modulare in cui tre reti 'esperte' (i moduli) erano in competizione per rispondere a immagini proiettate su una retina artificiale 5 × 5 (Fig. 4.10).

Un'unità specificava se si doveva eseguire il compito di identificazione o quello di localizzazione. La specificazione del compito era inviata alle tre reti esperte e a una rete di 'selezionamento' (*gating*), la quale sviluppava tre pesi (g1-g3) in grado di regolare la misura in cui ognuna delle tre reti esperte influenzava la risposta complessiva del sistema. Le reti esperte avevano differenti strutture: la prima era una rete a due strati (priva cioè di unità nascoste); le altre due avevano rispettivamente 18 e 36 unità nascoste. Dopo l'addestramento il sistema aveva appreso a suddividere i compiti in modo tale che la rete priva di unità nascoste eseguiva il compito 'dove' e la rete con 36 unità nascoste eseguiva il compito 'che cosa'. Questa

è in effetti la distribuzione ottimale dal momento che il compito di localizzazione è *linearmente separabile*, e dunque gestibile in modo efficiente da una rete priva di unità nascoste; mentre il compito di identificazione, più complesso, richiede unità nascoste. I ricercatori hanno confrontato questa prestazione con quella di una rete non modulare, constatando che la seconda, oltre a essere vulnerabile all'*interferenza catastrofica*[2], non è mai riuscita a raggiungere il livello di apprendimento della prima. Ciò induce a ipotizzare che la scomposizione del compito posta in essere dal sistema nervoso sia la più efficiente sotto il profilo computazionale.

3.2. *Funzionalismo senza realizzabilità multipla*

Il rapporto *dialettico* tra psicologia e neuroscienza – da un lato la conoscenza funzionale ottenuta grazie all'indagine psicologica consente l'identificazione dei meccanismi neuronali; dall'altro la conoscenza della struttura cerebrale offre una guida euristica per l'elaborazione di modelli psicologici più sofisticati – ha una conseguenza importante: la confutazione degli argomenti in favore della possibilità *empirica* (e non meramente logica) della realizzabilità multipla. Infatti la tesi della realizzabilità multipla raffigura le tassonomie della psicologia e della neuroscienza come pienamente autonome. Ma se questo corrispondesse a realtà, dovrebbe essere possibile costruire la tassonomia strutturale indipendentemente dalle scomposizioni funzionali della psicologia cognitiva. E inoltre la cartografia cerebrale non potrebbe costituire una guida attendibile per la scomposizione della funzione psicologica. L'esatto contrario, insomma, di quanto si constata analizzando le pratiche epistemiche dei neuroscienziati cognitivi.

Queste considerazioni preludono al divorzio tra funzionalismo e realizzabilità multipla. Il funzionalismo è innanzitutto una tesi *metafisica* (gli stati mentali *sono* ruoli causali),

[2] Si parla di 'interferenza catastrofica' quando l'apprendimento di un nuovo compito interferisce con la capacità della rete di eseguire un compito già appreso.

scaturita dal rifiuto della teoria dell'identità di tipo alla luce dell'argomento della realizzabilità multipla. Nel cognitivismo classico questa tesi metafisica si è legata al funzionalismo come strategia esplicativa: il momento della *scomposizione funzionale* nella costruzione di un modello meccanicista di un determinato processo cognitivo ('analisi funzionale' nella terminologia di Block 1980). Tuttavia in questo secondo senso il funzionalismo ha rappresentato per alcuni studiosi non una confutazione bensì una *conferma* della teoria dell'identità di tipo. Come osserva correttamente Nannini, la teoria causale della mente di David K. Lewis e il materialismo dello stato centrale di David M. Armstrong presentano rispetto al funzionalismo metafisico una differenza cruciale:

> Lewis e Armstrong non pensano che, descrivendo gli stati mentali mediante i ruoli causali da essi ricoperti, si sia colta la loro natura; pensano soltanto che sia stato compiuto un passo preliminare per poter più facilmente riconoscere tale natura. Essi non sono, insomma, dei funzionalisti [metafisici], ma dei materialisti: non ritengono che gli stati mentali siano semplicemente stati funzionali (che hanno certo bisogno di essere implementati da qualcosa, ma sono essenzialmente distinti dal loro supporto); pensano viceversa che gli stati mentali siano intrinsecamente ed essenzialmente degli stati cerebrali, che possono essere più facilmente (o anche esclusivamente) individuati mediante le funzioni che sono capaci di svolgere (Nannini 2002, p. 113).

In questa prospettiva, l'idea che gli stati psicologici sono *individuati* funzionalmente è sostenuta indipendentemente dalla tesi della realizzabilità multipla. Una volta separate le due tesi, diviene possibile combinare analisi funzionale e teoria dell'identità di tipo, sostenendo che i processi psicologici sono individuati in base alle loro interazioni, pur riconoscendo che sono identici a processi cerebrali. Questa è la soluzione del problema mente-cervello implicita nella dialettica della scomposizione e della localizzazione all'opera in neuroscienza cognitiva. Il momento della scomposizione corrisponde alla focalizzazione funzionalista sull'individuazione dei processi psicologici in base alle relazioni con altri processi psicologici; il momento della localizzazione fa sua la tesi identitista secondo cui i processi psicologici sono processi neurologici che hanno luogo in specifiche aree cerebrali (Bechtel *et al.* 2001, p. 15).

Queste considerazioni sono state organicamente sviluppate da McCauley e Bechtel (2001) in una variante *euristica* della teoria dell'identità di tipo. Fondata sulla ricostruzione delle pratiche dei neuroscienziati cognitivi, questa teoria interpreta le tesi di identità tra fenomeni psicologici e fenomeni cerebrali non già come l'esito della ricerca, bensì come ipotesi che guidano gli sviluppi dell'indagine. Ponendo in rilievo la natura integralmente *ipotetica* delle tesi di identità nella scienza, la teoria dell'identità come euristica della scoperta si focalizza sui modi in cui le corrispondenze interteoriche istituite tra fenomeni psicologici e fenomeni neurologici promuovono l'integrazione e l'avanzamento delle conoscenze in psicologia e in neuroscienza. Le identità ipotizzate promuovono la ricerca additando nuove vie all'indagine empirica sulla mente e il cervello. Le scoperte che ne conseguono inducono neuroscienziati e psicologi a rielaborare le proprie teorie dei processi e delle strutture cerebrali e psicologiche. L'evoluzione di queste identità ipotetiche è determinata dal divenire della ricerca. Sono i successi esplicativi e predittivi a fornire una giustificazione pragmatica di queste tesi di identità e a rendere disponibili nuove risorse teoriche e probative nella ricerca futura.

Contro l'obiezione della realizzabilità multipla, la teoria dell'identità come euristica della scoperta richiama l'attenzione sul contributo che, nel corso del tempo, le tesi di identità psiconeuronale hanno apportato ai programmi di ricerca 'progressivi' in neuroscienza e in psicologia. Dal punto di vista di una filosofia della mente 'orientata empiricamente' è difficile negare che tali considerazioni costituiscano un potente argomento in favore della teoria dell'identità di tipo.

Cos'altro leggere

1. *Il progetto computazionale in psicologia*

La tesi della sostanziale continuità tra il rappresentazionalismo empirista e il cognitivismo classico è avanzata da Fodor (2000, *Introduzione*, nota 5). Un'eccellente introduzione allo studio della *psicologia del senso comune* in filosofia e in psicologia è Meini (2001). Di Francesco (2000) consente di orientarsi nel complesso dibattito sulla *coscienza fenomenica*. Gozzano (1999) ricostruisce la storia del problema dell'*intenzionalità* da Brentano a Dennett. Von Eckardt (1994) offre una visione d'assieme delle varie prospettive sullo *statuto epistemico della psicologia del senso comune*.

La storia del concetto di *introspezione* è ricostruita da Lyons (1986). Una buona introduzione storica e teorica al *comportamentismo* è Mackenzie (1977). Il *comportamentismo eliminativista* è discusso con precisione filosofica da Byrne (1994). Un'utile visione d'assieme del *materialismo eliminativo* è Stich (1999). Smith (1986) esamina i rapporti tra il comportamentismo psicologico e quello logico.

Il lavoro di Tolman sulle *mappe cognitive* è all'origine degli attuali studi sulla navigazione spaziale, su cui vedi Gallistel (1999). Un'analisi dettagliata dei *limiti dello schema stimolo-risposta* in relazione agli esperimenti di Tolman e al programma della linguistica chomskiana è svolta da Rey (1997, cap. 4). L'*argomento della povertà dello stimolo* è ricostruito e difeso da varie critiche filosofiche in Laurence e

Margolis (2001). Bechtel *et al.* (1998) è un'accurata storia della *scienza cognitiva*. Esistono vari manuali di *psicologia cognitiva*: un buon punto di partenza è Eysenck e Keane (1990). Marconi (2001) è una splendida prima introduzione ai *rapporti tra filosofia e scienza cognitiva*. Due manuali di *filosofia della scienza cognitiva* sono Clark (2001) e Marraffa (2002).

Jervis (1999, p. 68) sviluppa il tema del *decentramento del soggetto*, individuandone in Jean Piaget il referente storico principale. Con quest'ultimo infatti «inizia a farsi strada la necessità di studiare la coscienza del bambino 'dal basso' e non già 'dall'alto', cioè cercando di formulare ipotesi che non fossero ricalcate sulle categorie fenomenologiche dell'autocoscienza adulta» (Jervis 1993, p. 201). Searle afferma e Dennett nega l'idea della *primarietà della soggettività cosciente* in Searle (1997, cap. 5).

Il *funzionalismo* fa la parte del leone in quasi tutti i manuali di filosofia della mente: in italiano si possono consultare Nannini (2001), Di Francesco (2002) e Paternoster (2002). Harel (2000) è un'introduzione elementare ai concetti fondamentali della *teoria della computazione*.

Quine (1960) argomenta in favore dell'*antirealismo intenzionale*. La storia del problema della *meccanizzazione della razionalità* è ben ricostruita da Haugeland (1985, cap. 1). Le difficoltà poste dal *problema dell'homunculus* allo psicologo sono esposte da R. Wilson (1999, pp. XVIII-XXI).

L'unica ricostruzione complessiva dell'opera di Jerry Fodor è Cain (2002). Fodor (1994) fornisce una brillante prima introduzione alla *teoria computazional-rappresentazionale della mente*. Esposizioni più avanzate sono Sterelny (1990), Rey (1997) e Aydede (1998). Aizawa (2003) è una puntuale rassegna del dibattito sulla *sistematicità* della cognizione. Paternoster (2002, cap. 4) e Adams (2003) discutono lucidamente il problema della *semantica delle rappresentazioni mentali*. Sul *problema della disgiunzione* vedi Paternoster (2001, par. 2.3). Una ricostruzione dell'evoluzione delle idee di Fodor in merito alla *semantica del mentalese* è contenuta in Ferretti (2001).

De Grada e Mannetti (1992) è un'approfondita rassegna sull'*attribuzione causale*. Gazzaniga (1985) ricostruisce lo sviluppo delle ricerche sulla *sindrome da disconnessione interemisferica*. T. Wilson (2002) estende le sue indagini sull'*inconscio cognitivo*. Stich dopo aver espresso i suoi dubbi eliminativisti (Stich 1983), è diventato uno dei principali critici dell'eliminativismo (Stich 1996).

2. *La spiegazione psicologica e la struttura della mente*

Su *leggi rigorose* e *clausole ceteris paribus*, vedi il classico Davidson (1970). L'opera di riferimento sulla *spiegazione meccanicista* è Bechtel e Richardson (1993). Vedi anche Machamer *et al.* (2000). Bechtel (1997) discute i differenti ruoli che possono ricoprire le nozioni di *legge* e di *meccanismo* in scienza cognitiva.

Bechtel *et al.* (2001, parte II) contiene alcuni testi classici della *neuropsicologia del linguaggio* (da Broca agli attuali studi di neuroimmagine). Il saggio finale di Bechtel introduce alla riflessione filosofica su questo ambito di indagine.

Cummins (1983) è il testo canonico sull'*analisi funzionale* in psicologia cognitiva. Boccignone (2001) e Paternoster (2001, cap. 5) sono due buone introduzioni alla *teoria della visione di Marr*. Papagno (2003) e Schacter (1996) ci guidano nello studio dell'*architettura della memoria*. Il *funzionalismo omuncolare* è stato proposto per la prima volta da Fodor (1968) e successivamente sviluppato da Dennett (1978, parte III) e Lycan (1981).

Chomsky (1988) è una presentazione elementare della teoria dei *principi e parametri*. Buone visioni d'assieme della *teoria della modularità* sono fornite da Garfield (1994), Carston (1996), Cain (2002, cap. 7). La distinzione tra modularità *sincronica* e modularità *diacronica* è tracciata da Segal (1996), mentre quella tra moduli *chomskiani, computazionali* e *chomskiano-computazionali* è sviluppata da Samuels (1998; 2000). Sull'incapsulamento informativo della *percezione visiva* è fondamentale la lettura di Pylyshyn (1999). Flores d'Arcais (1993) è un'introduzione alla psicolinguistica che presenta accuratamente gli studi sul *riconoscimento di parole* (cap. 7) e sull'*ela-*

borazione sintattica nella comprensione linguistica (cap. 8). Per un punto di vista sul riconoscimento lessicale diverso da quello di Swinney, vedi Tabossi e Zardon (1993).

Haselager (1997) è molto utile per orientarsi nel dibattito sull'*olismo della cognizione centrale*. Samuels (1998; 2000) fornisce un'analisi accurata dell'*ipotesi della modularità massiva*. Una critica sistematica di tale ipotesi è invece contenuta in Fodor (2000). L'ipotesi è solo una delle strategie di risposta al problema dell'olismo cognitivo. Waskan e Bechtel (1997) e Carruthers (2003) contestano la legittimità del *passaggio dall'olismo epistemologico all'olismo cognitivo* compiuto da Fodor; mentre Horgan e Tienson (1996) giudicano irresolubile il problema per il cognitivismo classico ed esplorano la possibilità di sostituire il formalismo computazionale con quello della teoria dei sistemi dinamici.

3. *Psicologia ingenua e razionalità naturale*

La distinzione tra *i due sensi in cui la psicologia del senso comune può venir concepita come una teoria* – come teoria implicita nel discorso ordinario sugli stati mentali o come teoria rappresentata internamente – è tracciata da Stich e Ravenscroft (1994).

Tre volumi collettanei curati da Davies e Stone (1995a; 1995b), e da Carruthers e Smith (1996) consentono di farsi un'idea del *dibattito interdisciplinare sulla psicologia ingenua*. Camaioni (2003) mette a disposizione del lettore italiano alcuni classici contributi alla *Teoria della Mente*. Il *compito della credenza falsa* è stato messo a punto da Wimmer e Perner (1983). Un'utile rassegna dei risultati ottenuti impiegando questo paradigma sperimentale è Wellman *et al.* (2001). I testi di David e Cutting (1994), Grivois e Proust (1998), Coltheart e Davies (2000), Surian (2002), Ferretti (2003), discutono le prospettive dell'approccio metarappresentazionale in relazione all'autismo, la schizofrenia, le sindromi da identificazione erronea, i disturbi di personalità e altre patologie della mente.

Il manifesto della *psicologia evoluzionistica* è il volume curato da Barkow *et al.* (1992). La più recente introduzione a questo controverso programma di ricerca è Buss (1999).

Girotto (1994) è una chiara introduzione alla *psicologia del ragionamento*. Stein (1996) è un'eccellente ricostruzione del *dibattito sulla razionalità* in filosofia e in scienza cognitiva. Le conseguenze della psicologia del ragionamento per il dibattito sulla *razionalità naturale* sono indagate da Samuels *et al.* (1999) e Samuels *et al.* (2003), che proseguono l'opera iniziata da Stich (1991).

4. *Psicologia e neuroscienza*

Una presentazione della *teoria dell'identità di tipo*, estremamente autorevole in quanto scritta da uno dei suoi padri fondatori, è Smart (2000). Sull'importanza del modello della *riduzione interteorica* per la filosofia della mente contemporanea è indispensabile la lettura di Bickle (2001).

Una rassegna aggiornata sulla tesi della *realizzabilità multipla* è Bickle (1997); una discussione più tecnica e dettagliata di alcuni temi della rassegna si può trovare in Bickle (1998). Il dibattito sulla tesi dell'*autonomia della psicologia* è ben ricostruito in Bechtel (1988, cap. 4).

Sull'ascesa della *neuroscienza cognitiva* e la sua centralità nell'odierna scienza cognitiva, vedi Bechtel *et al.* (1998, pp. 82-86). Per un primo accostamento al *connessionismo*, vedi Frixione (1992), Paternoster (2002, capp. 6-7) e Marraffa (2002, cap. 4). Un'esposizione più avanzata è Bechtel e Abrahamsen (2001). Questo testo (pp. 67-69) contiene anche una chiara spiegazione del concetto di separabilità lineare a cui si è fatto riferimento a p. 135. Macdonald e Macdonald (1995) raccoglie alcuni tra i più importanti contributi alla riflessione filosofica sul connessionismo. McCauley (1996) esamina criticamente il modello della coevoluzione eliminativista e riduzionista.

Sul pluralismo esplicativo, oltre al fascicolo di «Theory and Psychology» (vol. 11, 2001) che ospita vari interventi sul tema (incluso quello di McCauley e Bechtel discusso nel testo), vedi Bechtel e Richardson (1993), Flanagan (1992), Hardcastle (1996), McCauley (1996), Looren de Jong (1997).

Bibliografia

Adams, F. (2003), *Thoughts and their contents: naturalized semantics*, in S. Stich e T. Warfield (a cura di), *The Blackwell Guide to Philosophy of Mind*, Blackwell, Oxford.

Aizawa, K. (2003), *Cognitive architecture: the structure of cognitive representations*, in S. Stich e T. Warfield (a cura di), *The Blackwell Guide to Philosophy of Mind*, Blackwell, Oxford.

Atkinson, R., Shiffrin, R. (1968), *Human memory: a proposed system and its control processes*, in K. Spence e J. Spence (a cura di), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Academic Press, New York, vol. II.

Aydede, M. (1998), *Language of thought hypothesis*, in E. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy*, URL: *http://plato.stanford.edu/entries/language-thought*

Baddeley, A., Hitch, J. (1974), *Working memory*, in G.H. Bower, (a cura di), *The Psychology of Learning and Motivation*, Academic Press, New York, vol. VIII.

Barkow, J., Cosmides, L., Tooby, J. (1992) (a cura di), *The Adapted Mind. Evolutionary Psychology and the Generation of Culture*, Oxford University Press, Oxford.

Baron-Cohen, S. (1995), *Mindblindness*, MIT Press, Cambridge (MA).

Baron-Cohen, S., Leslie, A., Frith, U. (1985), *Does the autistic child have a 'theory of mind'?*, in «Cognition», 21, pp. 37-46.

Bechtel, W. (1988), *Philosophy of Science. An Overview for Cognitive Science*, Erlbaum, Hillsdale (NJ) [trad. it. *Filosofia della scienza e scienza cognitiva*, Laterza, Roma-Bari 2001²].

Bechtel, W. (1997), *Representations and cognitive explanations: as-*

sessing the dynamicist's challenge in cognitive science, in «Cognitive Science», 22, pp. 295-318.
Bechtel, W., Abrahamsen, A. (2001), *Connectionism and the Mind: Parallel Processing, Dynamics, and Evolution in Networks*, Blackwell, Oxford.
Bechtel, W., Abrahamsen, A., Graham, G. (1998), *The life of cognitive science*, in W. Bechtel e G. Graham (a cura di), *A Companion to Cognitive Science*, Blackwell, Oxford, pp. 1-104 [trad. it. *Breve storia della scienza cognitiva*, Laterza, Roma-Bari 2004].
Bechtel, W., Mandik, P., Mundale, J., Stufflebeam, R.S. (a cura di) (2001), *Philosophy and the Neurosciences: A Reader*, Blackwell, Oxford.
Bechtel, W., Mundale, J. (1999), *Multiple realizability revisited: linking cognitive and neural states*, in «Philosophy of Science», 66, pp. 175-207.
Bechtel, W., Richardson, R.C. (1993), *Discovering Complexity*, Princeton University Press, Princeton.
Bermúdez, J. (2000), *Personal and subpersonal: a difference without a distinction*, in «Philosophical Explorations», 3, pp. 63-82.
Bickle, J. (1997), *Multiple realizability*, in E. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy*, URL: *http://plato.stanford.edu/entries/multiple-realizability*
Bickle, J. (1998), *Psychoneural Reduction: The New Wave*, MIT Press, Cambridge (MA).
Bickle, J. (2001), *Concepts of intertheoretic reduction in contemporary philosophy of mind*, in M. Marraffa e M. Nani (a cura di), *A Field Guide to the Philosophy of Mind*, URL: *http://www.host.uniroma3.it/kant/field/cir*
Bickle, J., Mandik, P. (2002), *The philosophy of neuroscience*, in E.N. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy*, URL: *http://plato.stanford.edu/archives/win2002/entries/neuroscience*
Block, N. (1980), *Introduction: what is functionalism?*, in N. Block (a cura di), *Readings in the Philosophy of Psychology*, Harvard University Press, Cambridge (MA), vol. I.
Block, N. (1995), *On a confusion about a function of consciousness*, in «Behavioral and Brain Sciences», 18, pp. 227-247.
Boccignone, G. (2001), *Visione artificiale*, in E. Burattini e R. Cordeschi (a cura di), *Intelligenza artificiale*, Carocci, Roma.
Botterill, G., Carruthers, P. (1999), *The Philosophy of Psychology*, Cambridge University Press, Cambridge [trad. it. *Filosofia della psicologia*, Il Saggiatore, Milano 2003].
Buss, D.M. (1999), *Evolutionary Psychology: The New Science of the Mind*, Allyn and Bacon, Boston.
Byrne, A. (1994), *Behaviourism*, in S. Guttenplan (a cura di), *A Companion to the Philosophy of Mind*, Blackwell, Oxford.
Cain, M.J. (2002), *Fodor. Language, Mind and Philosophy*, Polity Press, Cambridge.
Camaioni, L. (a cura di) (2003), *Teoria della mente*, Laterza, Roma-Bari.
Carey, S., Spelke, E. (1996), *Science and core knowledge*, in «Philosophy of Science», 63, pp. 515-533.
Carruthers, P. (2003), *Moderately massive modularity*, in A.O'Hear (a cura di), *Mind and Persons*, Cambridge University Press, Cambridge.
Carruthers, P., Smith, P. (1996), *Theories of Theories of Mind*, Cambridge University Press, Cambridge.
Carston, R. (1996), *The architecture of the mind: modularity and modularization*, in D.W. Green *et al.*, *Cognitive Science. An Introduction*, Blackwell, Oxford.
Cartwright, N. (1983), *How the Laws of Physics Lie*, Oxford University Press, Oxford.
Chalmers, D. (1996), *The Conscious Mind*, Oxford University Press, Oxford [trad. it. *La mente cosciente*, McGraw-Hill, Milano 1999].
Cheng, P., Holyoak, K. (1985), *Pragmatic reasoning schemas*, in «Cognitive Psychology», 17, pp. 391-416.
Chomsky, N. (1988), *Language and Problems of Knowledge. The Managua Lectures*, MIT Press, Cambridge (MA) [trad. it. *Linguaggio e problemi della conoscenza*, Il Mulino, Bologna 1998].
Churchland, P.M. (1979), *Scientific Reason and the Plasticity of Mind*, Cambridge University Press, Cambridge.
Churchland, P.M. (1981), *Eliminative materialism and the propositional attitudes*, in «Journal of Philosophy», 78, pp. 67-90; rist. in Churchland (1989).
Churchland, P.M. (1989), *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, MIT Press, Cambridge (MA) [trad. it. parz. *La natura della mente e la struttura della scienza*, Il Mulino, Bologna 1992].
Churchland, P.M. (1990), *Cognitive activity in artificial neural*

networks, in D. Osherson e E.E. Smith (a cura di), *An Invitation to Cognitive Science*, MIT Press, Cambridge (MA), vol. 3.
Churchland, P.M. (1998), *Evaluating our self-conception*, in P.M. Churchland e P.S. Churchland, *On the Contrary*, MIT Press, Cambridge (MA).
Churchland, P.M., Churchland, P.S. (1996), *McCauley's demand for a co-level competitor*, in R.N. McCauley (a cura di), *The Churchlands and their Critics*, Blackwell, Oxford.
Churchland, P.M., Churchland, P.S. (1998), *On the Contrary*, MIT Press, Cambridge (MA).
Churchland, P.S. (1986), *Neurophilosophy*, MIT Press, Cambridge (MA).
Clark, A. (2001), *Mindware. An Introduction to the Philosophy of Cognitive Science*, Oxford University Press, Oxford.
Coltheart, M., Davies, M. (2000) (a cura di), *Pathologies of Belief*, Blackwell, Oxford.
Cosmides, L. (1989), *The logic of social exchange: has natural selection shaped how humans reason? Studies with Wason Selection Task*, in «Cognition», 31, pp. 187-276.
Cosmides, L., Tooby, J. (1992), *Cognitive adaptations for social exchange*, in J. Barkow, L. Cosmides e J. Tooby (a cura di), *The Adapted Mind*. Oxford University Press, Oxford.
Cosmides, L., Tooby, J. (1996), *Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty*, in «Cognition», 58(1), pp. 1-73.
Cummins, R. (1983), *The Nature of Psychological Explanation*, MIT Press, Cambridge (MA).
Darden, L., Maull, N. (1987), *Interfield theories*, in «Philosophy of Science», 43, pp. 44-64.
David, A., Cutting, J. (1994) (a cura di), *The Neuropsychology of Schizophrenia*, Erlbaum, Hillsdale (NJ).
Davidson, D. (1970), *Mental events*, in L. Forster e J.W. Swanson (a cura di), *Experience and Theory*, University of Massachusetts Press, Amherst (MA), pp. 79-101 [trad. it. *Eventi mentali*, in Id., *Azioni ed eventi*, Il Mulino, Bologna 1992, pp. 285-309].
Davies M., Stone T. (1995a) (a cura di), *Folk Psychology*, Blackwell, Oxford.
Davies M., Stone T. (1995b) (a cura di), *Mental Simulation*, Blackwell, Oxford.

De Grada, E., Mannetti, L. (1992), *L'attribuzione causale*, Il Mulino, Bologna.
Dennett, D. (1969), *Content and Consciousness*, Routledge, London [trad. it. *Contenuto e coscienza*, Il Mulino, Bologna 1992].
Dennett, D. (1978), *Brainstorms*, MIT Press, Cambridge (MA) [trad. it. *Brainstorms*, Adelphi, Milano 1991].
Dennett, D. (1993), *Review of John Searle, "The Rediscovery of the Mind"*, in «The Journal of Philosophy», 90, pp. 193-205.
Di Francesco, M. (2000), *La coscienza*, Laterza, Roma-Bari.
Di Francesco, M. (2002), *Introduzione alla filosofia della mente*, Carocci, Roma.
Eysenck, M., Keane, M. (1990), *Cognitive Psychology. A Student's Handbook*, Erlbaum, Hillsdale (NJ) [trad. it. *Manuale di psicologia cognitiva*, Sorbona, Milano 1995].
Ferreira, F., Clifton, C. (1986), *The independence of syntactic processing*, in «The Journal of Memory and Language», 25, pp. 348-368.
Ferretti, F. (2001), *Naturalizzare le menti: un percorso introduttivo al pensiero di Jerry A. Fodor*, in J. Fodor, *Mente e linguaggio*, Laterza, Roma-Bari.
Ferretti, F. (2003) (a cura di), *La mente dell'altro. Prospettive sull'autismo*, Editori Riuniti, Roma.
Fiedler, K. (1988), *The dependence of the conjunction fallacy on subtle linguistic factors*, in «Psychological Research», 50, pp. 123-129.
Flanagan, O. (1992), *Consciousness Reconsidered*, MIT Press, Cambridge (MA).
Flores d'Arcais, G.B. (1993), *La psicolinguistica*, Cleup, Padova.
Fodor, J.A. (1968), *The appeal to tacit knowledge in psychological explanation*, in «Journal of Philosophy», 65, pp. 627-640 [rist. in Id., *Representations*, MIT Press, Cambridge (MA)].
Fodor, J.A. (1975), *The Language of Thought*, Crowell, New York.
Fodor, J.A. (1983), *The Modularity of Mind*, MIT Press, Cambridge (MA) [trad. it. *La mente modulare*, Il Mulino, Bologna 1988].
Fodor, J.A. (1986), *Banish DisContent*, in J. Butterfield (a cura di), *Language, Mind and Logic*, Cambridge University Press, Cambridge.
Fodor, J.A. (1987), *Psychosemantics*, MIT Press, Cambridge (MA) [trad. it. *Psicosemantica*, Il Mulino, Bologna 1990].
Fodor, J.A. (1991), *Too hard for our kind of mind?*, in «London Review of Books», 27 giugno.

Fodor, J.A. (1994), *Fodor, Jerry A.*, in S. Guttenplan (a cura di), *A Companion to the Philosophy of Mind*, Blackwell, Oxford.
Fodor, J.A. (2000), *The Mind Doesn't Work That Way*, MIT Press, Cambridge (MA) [trad. it. *La mente non funziona così*, Laterza, Roma-Bari 2001].
Fodor, J.A. (2001), *Mente e linguaggio*, a cura di F. Ferretti, Laterza, Roma-Bari.
Fodor, J.A., Pylyshyn, Z. (1988), *Connectionism and cognitive architecture: A critical analysis*, in «Cognition», 28, pp. 3-71; rist. in Macdonald e Macdonald (1995).
Frazier, L. (1978), *On Comprehending Sentences: Syntactic Parsing Strategies*, Indiana University Linguistics Pub., Bloomington (IN).
Frazier, L., Rayner, K. (1982), *Making and correcting errors during sentence comprehension: eye movements in the analysis of structurally ambiguous sentences*, in «Cognitive Psychology», 14, pp. 178-210.
Frixione, M. (1992), *I modelli della mente tra simboli e neuroni: il connessionismo subsimbolico di Smolensky*, in P. Smolensky *Il connessionismo. Tra simboli e neuroni*, Marietti, Genova.
Gallistel, C.R. (1999), *Animal navigation*, in R. Wilson e F. Keil (a cura di), *The MIT Encyclopedia of the Cognitive Sciences*, MIT Press, Cambridge (MA).
Garfield, J.L. (1994), *Modularity*, in S. Guttenplan (a cura di), *A Companion to the Philosophy of Mind*, Blackwell, Oxford.
Gazzaniga, M. (1985), *The Social Brain*, Basic Books, New York [trad. it. *Il cervello sociale*, Giunti, Firenze 1989].
Gazzaniga, M. (1991), *La plasticità del sistema nervoso centrale e il ruolo dell'interprete*, in G. Giorello e P. Strata (a cura di), *L'automa spirituale*, Laterza, Roma-Bari.
Geschwind, N. (1979), *Specializations of the human brain*, in «Scientific American», 238(3), pp. 158-168.
Gigerenzer, G. (1994), *Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa)*, in G. Wright e P. Ayton (a cura di), *Subjective Probability*, John Wiley, New York.
Girotto, V. (1994), *Il ragionamento*, Il Mulino, Bologna.
Gödel, K. (1930), *Der Vollständigkeit der Axiome des logischen Funktionenkalküls*, in «Monatshefte für Mathematik und Physik», 37, pp. 349-360.

Goldman, A. (1993), *The psychology of folk psychology*, in «Behavioral and Brain Sciences», 16, pp. 15-28, 101-113.
Gopnik, A., Meltzoff, A. (1994), *Minds, bodies, and persons: young children's understanding of the self and others as reflected in imitation and theory of mind research*, in S. Parker, R. Mitchell e M. Boccia (a cura di), *Self-awareness in Animals and Humans*, Cambridge University Press, Cambridge.
Gopnik, A., Meltzoff, A. (1997), *Words, Thoughts, and Theories*, MIT Press, Cambridge (MA) [trad. it. *Costruire il mondo*, McGraw-Hill, Milano 2001].
Gorman, R., Sejnowski, T. (1988), *Learned classification of sonar targets using a massively parallel network*, in «IEEE Transactions: Acoustics, Speech and Signal Processing», 36, pp. 1135-1140.
Gozzano, S. (1999), *Storia e teorie dell'intenzionalità*, Laterza, Roma-Bari.
Gregory, R.L. (1997), *Eye and Brain. The Psychology of Seeing*, Oxford University Press, Oxford (V ed.).
Grivois, H., Proust, J. (1998), *Subjectivité et conscience d'agir*, PUF, Paris.
Hardcastle, V. (1996), *How to Build a Theory in Cognitive Science*, John Benjamins, Philadelphia (PN).
Harel, D. (2000), *Computers Ltd. What They Really Can't Do*, Oxford University Press, Oxford.
Haselager, W.F.G. (1997), *Cognitive Science and Folk Psychology*, Sage, London.
Haugeland, J. (1985), *Artificial Intelligence: The Very Idea*, MIT Press, Cambridge (MA) [trad. it. *Intelligenza artificiale*, Bollati Boringhieri, Torino 1988].
Heider, F. (1958), *The Psychology of Interpersonal Relations*, John Wiley, New York [trad. it. *Psicologia delle relazioni interpersonali*, Il Mulino, Bologna 1972].
Heider, F., Simmel, M. (1944), *An experimental study of apparent behavior*, in «American Journal of Psychology», 57, pp. 243-259.
Horgan, T., Tienson, J. (1996), *Connectionism and the Philosophy of Psychology*, MIT Press, Cambridge (MA).
Horst, S. (1996), *Symbols, Computation and Intentionality*, University of California Press, Berkeley (CA).
Hull, C. (1943), *Principles of Behavior*, Appleton-Century-Crofts, New York.

Jacobs, R.A., Jordan, M.I., Barto, A.G. (1991), *Task decomposition through competition in a modular connectionist architecture: the what and where vision tasks*, in «Cognitive Science», 15, pp. 219-250.
Jacoby, H. (1985), *Eliminativism, meaning and qualitative states*, in «Philosophical Studies», 47, pp. 257-270.
Jervis, G. (1993), *Fondamenti di psicologia dinamica*, Feltrinelli, Milano.
Jervis, G. (1999), *Prime lezioni di psicologia*, Laterza, Roma-Bari.
Kahneman, D., Tversky, A. (1973), *On the psychology of prediction*, in «Psychological Review», 80, pp. 237-251.
Kuhn, T. (1962), *The Structure of Scientific Revolutions*, University of Chicago Press, Chicago.
Latané, B., Darley, J.M. (1970), *The Unresponsive Bystander: Why Doesn't He Help?*, Appleton-Century-Crofts, New York.
Laurence, S., Margolis, E. (2001), *The poverty of the stimulus argument*, in «The British Journal for the Philosophy of Science», 52, pp. 217-276.
Leslie, A. (1987), *Pretense and representation: the origins of 'theory of mind'*, in «Psychological Review», 94, pp. 412-426.
Leslie, A. (1994), *ToMM, ToBy, and agency: core architecture and domain specificity*, in L. Hirschfeld e S. Gelman (a cura di), *Mapping the mind: Domain specificity in Cognition and Culture*, Cambridge University Press, Cambridge (MA).
Leslie, A. (2000), *'Theory of Mind' as a mechanism of selective attention*, in M.S. Gazzaniga (a cura di), *The New Cognitive Neurosciences*, MIT Press, Cambridge (MA), pp. 1235-1247.
Leslie, A., Thaiss, L. (1992), *Domain specificity in conceptual development: neuropsychological evidence from autism*, in «Cognition», 43, pp. 225-251.
Lewis, D.K. (1972), *Psychophysical and theoretical identifications*, in «Australasian Journal of Philosophy», 50, pp. 249-58.
Loar, B. (1983), *Must beliefs be sentences?* In P. Asquith e T. Nickles (a cura di), *PSA 1982. Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association*, Philosophy of Science Association, East Lansing (MI), vol. 2.
Looren de Jong, H. (1997), *Levels: reduction and elimination in cognitive neuroscience*, in C.W. Tolman, F. Cherry, R. van Hezewijk, I. Lubek (a cura di), *Problems of Theoretical Psychology*, Captus Press, New York.

Lycan, W. (1981), *Form, function and feel*, in «Journal of Philosophy», 78, pp. 24-50.
Lycan, W. (1988), *Judgement and Justification*, Cambridge University Press, Cambridge.
Lycan, W. (1994), *Functionalism (1)*, in S. Guttenplan (a cura di), *A Companion to the Philosophy of Mind*, Blackwell, Oxford.
Lyons, W. (1986), *The Disappearance of Introspection*, MIT Press, Cambridge (MA) [trad. it. *La scomparsa dell'introspezione*, Il Mulino, Bologna 1993].
Macdonald, C., Macdonald, G. (a cura di) (1995), *Connectionism. Debates on Psychological Explanation*, Blackwell, Oxford.
Machamer, P., Darden L., Craver C. (2000), *Thinking about mechanisms*, in «Philosophy of Science», 67, pp. 1-25.
Mackenzie, B.D. (1977), *Behaviourism and the Limits of Scientific Method*, Routledge, London [trad. it. *Il comportamentismo e i limiti del metodo scientifico*, Armando, Roma 1980].
Marconi, D. (2001), *Filosofia e scienza cognitiva*, Laterza, Roma-Bari.
Marr, D. (1982), *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*, W.H. Freeman, San Francisco.
Marr, D., Nishihara, H.K. (1978), *Representation and recognition of the spatial organization of three dimensional structure*, in «Proceedings of the Royal Society of London B», 200, pp. 269-294.
Marraffa, M. (2002), *Scienza cognitiva. Un'introduzione filosofica*, Cleup, Padova.
Mazzoni, G. (2000), *L'apprendimento*, Carocci, Roma.
McCauley, R.N. (1996), *Explanatory pluralism and the coevolution of theories in science*, in R.N. McCauley (a cura di), *The Churchlands and their Critics*, Blackwell, Oxford.
McCauley, R., Bechtel, W. (2001), *Explanatory pluralism and the heuristic identity theory*, in «Theory and Psychology», 11, pp. 736-760.
Meini, C. (2001), *La psicologia ingenua*, McGraw-Hill, Milano.
Mishkin, M., Ungerleider, L.G., Macko, K.A. (1983), *Object vision and spatial vision: two cortical pathways*, in «Trends in Neurosciences», 6, pp. 414-417.
Morton, A. (1980), *Frames of Mind*, Oxford University Press, Oxford.
Murphy, D., Stich, S. (2000), *Darwin in the madhouse: evolutionary psychology and the classification of mental disorders*, in P. Car-

ruthers e A. Chamberlain (a cura di), *Evolution and the Human Mind*, Cambridge University Press, Cambridge.
Nagel, E. (1961), *The Structure of Science*, Harcourt, Brace, and World, New York.
Nannini, S. (2002), *L'anima e il corpo. Un'introduzione storica alla filosofia della mente*, Laterza, Roma-Bari.
Neisser, U. (1967), *Cognitive Psychology*, Appleton, New York [trad. it. *Psicologia cognitivista*, Martello-Giunti, Firenze 1976].
Nichols, S., Stich, S. (2001), *Leggere la propria mente*, in «Sistemi Intelligenti», XIII(1), pp. 143-170.
Nichols, S., Stich, S. (2002), *Reading one's own mind: a cognitive theory of self-awareness*, in Q. Smith e A. Jokic (a cura di), *Aspects of Consciousness*, Oxford University Press, Oxford.
Nisbett, R., Wilson, T. (1977), *Telling more than we can know: verbal reports on mental processes*, in «Psychological Review», 84(3), pp. 231-259.
O' Brien, G. (1993), *The connectionist vindication of folk psychology*, in S. Christensen e D. Turner (a cura di), *Folk Psychology*, Erlbaum, Hillsdale (NJ).
O'Keefe, J., Nadel, L. (1978), *The Hippocampus As A Cognitive Map*, Clarendon Press, Oxford.
Papagno, C. (2003), *Come funziona la memoria*, Laterza, Roma-Bari.
Paternoster, A. (2001), *Linguaggio e visione*, Edizioni ETS, Pisa.
Paternoster, A. (2002), *Introduzione alla filosofia della mente*, Laterza, Roma-Bari.
Perner, J. (1991), *Understanding the Representational Mind*, MIT Press, Cambridge (MA).
Perner, J. (1993), *The theory of mind deficit in autism: rethinking the metarepresentation theory*, in S. Baron-Cohen *et al.* (a cura di), *Understanding Other Minds: Perspectives from Autism*, Oxford University Press, Oxford.
Perner, J., Baker, S., Hutton, D. (1994), *Prelief: the conceptual origins of belief and pretence*, in C. Lewis e P. Mitchell (a cura di), *Children's Early Understanding of Mind*, Erlbaum, Hillsdale (NJ).
Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E. (1989), *Positron emission tomographic studies of the processing of single words*, in «Journal of Cognitive Neuroscience», 1, pp. 153-170.
Putnam, H. (1967), *Psychological predicates*, in W.H. Capitan e D.D. Merrill (a cura di), *Art, Mind, and Religion*, University of Pittsburgh Press, Pittsburgh (PA), pp. 37-48 [trad. it. *La natura degli stati mentali*, in Id., *Mente, linguaggio e realtà*, Adelphi, Milano 1987, pp. 461-473].
Pylyshyn, Z. (1991), *The role of cognitive architectures in theories of cognition*, in K. VanLehn (a cura di), *Architectures for Intelligence*, Erlbaum, Hillsdale (NJ).
Pylyshyn, Z. (1999), *Is vision continuous with cognition? The case for cognitive impenetrability of visual perception*, in «Behavioral and Brain Sciences», 22, pp. 341-365.
Quine, W.V. (1960), *Word and Object*, MIT Press, Cambridge (MA) [trad. it. *Parola e oggetto*, Il Saggiatore, Milano 1970].
Rayner, K. Carlson, M., Frazier, L. (1983), *The interaction of syntax and semantics during sentence comprehension: eye movements and the analysis of semantically biased sentences*, in «Journal of Verbal Learning and Verbal Behaviour», 22, pp. 358 –374.
Rey, G. (1997), *Contemporary Philosophy of Mind*, Blackwell, Oxford.
Rips, L.J. (1994), *The Psychology of Proof*, MIT Press, Cambridge (MA).
Rueckl, J.G., Cave, K.R., Kosslyn, S.M. (1989), *Why are "what" and "where" processed by separate cortical visual systems? A computational investigation*, in «Journal of Cognitive Neuroscience», 1, pp. 171-186.
Ruffman, T., Perner, J., Naito, M., Parkin, L., Clemens, W.A. (1998), *Older (but not younger) siblings facilitate false belief understanding*, in «Developmental Psychology», 34, pp. 161-174.
Rumelhart, D.E., McClelland, J.L. (1986), *PDP models and general topics in cognitive science*, in D.E. Rumelhart e J.L. McClelland (a cura di), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, MIT Press, Cambridge (MA), vol. I [trad. it. *Modelli PDP e temi generali della scienza cognitiva*, in D.E. Rumelhart e J.L. McClelland, *PDP. Microstruttura dei processi cognitivi*, Il Mulino, Bologna 1991].
Rumelhart, D.E., Smolensky, P., McClelland, J.L., Hinton, G.E. (1986), *Schemata and sequential thought processes in parallel distributed models*, in J.L. McClelland e D.E. Rumelhart (a cura di), *Parallel Distributed Processing. Explorations in the Microstructure of Cognition*, MIT Press, Cambridge (MA), vol. II [trad. it. *Schemi e processi di pensiero sequenziali nei modelli di*

tipo PDP, in D.E. Rumelhart e J.L. McClelland, *PDP. Microstruttura dei processi cognitivi*, Il Mulino, Bologna 1991].
Samuels, R. (1998), *Evolutionary psychology and the massive modularity hypothesis*, in «British Journal for the Philosophy of Science», 49, pp. 575-602.
Samuels, R. (2000), *Massively modular minds: evolutionary psychology and cognitive architecture*, in P. Carruthers e A. Chamberlain (a cura di), *Evolution and the Human Mind*, Cambridge University Press, Cambridge.
Samuels, R., Stich, S., Tremoulet, P. (1999), *Rethinking rationality: from bleak implications to darwinian modules*, in E. Lepore e Z. Pylyshyn (a cura di), *What is Cognitive Science?*, Blackwell, Oxford.
Samuels, R., Stich, S., Faucher, L. (2003), *Reason and rationality*, in M. Sintonen (a cura di), *Handbook of Epistemology*, Kluwer, Dordrecht.
Schacter, D.L. (1996), *Searching for Memory. The Brain, the Mind, and the Past*, Basic Books, New York [trad. it. *Alla ricerca della memoria*, Einaudi, Torino 2001].
Scholl, B.J., Leslie, A.M. (1999), *Modularity, development and 'theory of mind'*, in «Mind and Language», 14, pp. 131-153.
Searle, J.R. (1997), *The Mistery of Consciousness*, New York Review of Books, New York [trad. it. *Il mistero della coscienza*, Cortina, Milano 1998].
Segal, G. (1996). *The modularity of theory of mind*, in P. Carruthers e P. K. Smith, a cura di, *Theories of Theory of Mind*, Cambridge University Press, Cambridge (MA).
Seidenberg, M., Tannenhaus, M., Leiman, J., Bienkowski, M. (1982), *Automatic access of the meaning of ambiguous words in context: some limitation of knowledge-based processing*, in «Cognitive Psychology», 14, pp. 489-537.
Simon, H. (1969), *The Sciences of the Artificial*, MIT Press, Cambridge (MA) [trad. it. *Le scienze dell'artificiale*, Il Mulino, Bologna 1988].
Skinner (1984), *Skinner: canonical papers*, in «Behavioral and Brain Sciences», 7, pp. 473-701.
Slovic, P., Fischhoff, B., Lichtenstein, S. (1976), *Cognitive processes and societal risk taking*, in J.S. Carol e J.W. Payne (a cura di), *Cognition and Social Behavior*, Erlbaum, Hillsdale (NJ).
Smart, J.J.C. (2000), *The identity theory of mind*, in E. Zalta (a cura di), *The Stanford Encyclopedia of Philosophy*, URL: http://plato.stanford.edu/entries/mind-identity.
Smith, L.D. (1986), *Behaviorism and Logical Positivism: A Reassessment of the Alliance*, Stanford University Press, Stanford.
Sperber, D. (1994), *The Modularity of thought and the epidemiology of representations*, in L. Hirschfeld e S. Gelman (a cura di), *Mapping the mind: Domain specificity in cognition and culture*, Cambridge University Press, Cambridge (MA).
Sperber, D. (1996), *Explaining Culture. A Naturalistic Approach*, Blackwell, Oxford [trad. it. *Il contagio delle idee*, Feltrinelli, Milano 1999].
Stalnaker, R. (1984), *Inquiry*, MIT Press, Cambridge (MA).
Stein, E. (1996), *Without Good Reason. The Rationality Debate in Philosophy and Cognitive Science*, Clarendon Press, Oxford.
Sterelny, K. (1990), *The Representational Theory of Mind*, Blackwell, Oxford.
Stich, S. (1983), *From Folk Psychology to Cognitive Science*, MIT Press, Cambridge (MA) [trad. it. *Dalla psicologia del senso comune alla scienza cognitiva*, Il Mulino, Bologna, 1991].
Stich, S. (1991), *The Fragmentation of Reason*, MIT Press, Cambridge (MA) [trad. it. *La frammentazione della ragione*, Il Mulino, Bologna 1996].
Stich, S. (1996), *Deconstructing the Mind*, Oxford University Press, Oxford.
Stich, S. (1999), *Eliminative materialism*, in R. Wilson e F. Keil (a cura di), *The MIT Encyclopedia of the Cognitive Sciences*, MIT Press, Cambridge (MA).
Stich, S., Nichols, S. (2003), *Folk psychology*, in S. Stich e T. Warfield (a cura di), *The Blackwell Guide to Philosophy of Mind*, Blackwell, Oxford.
Stich, S., Ravenscroft, I. (1994), *What* is *folk psychology?*, in «Cognition», 50, pp. 447-468 [rist. in Stich 1996].
Stone, T., Young, A.W. (1997), *Delusions and brain injury: the philosophy and psychology of belief*, in «Mind and Language», 12, pp. 327-364.
Surian, L. (2002), *Autismo. Indagini sullo sviluppo mentale*, Laterza, Roma-Bari.
Swinney, D. (1979), *Lexical access during sentence comprehension: (re)consideration of context effects*, in «Journal of Verbal Learning and Verbal Behaviour», 2, pp. 219-227.

Tabossi, P., Zardon, F. (1993), *Processing ambiguous words in context*, in «Journal of Memory and Language», 32, pp. 359-372.
Tolman, E.C. (1948), *Cognitive maps in rats and men*, in «Psychological Review», 55, pp. 189-208.
Tversky, A., Kahneman, D. (1973), *Availability: a heuristic for judging frequency and probability*, in «Cognitive Psychology», 4, pp. 207-232.
Tversky, A., Kahneman, D. (1983), *Extensional versus intuitive reasoning: the conjunction fallacy in probability judgement*, in «Psychological Review», 90, pp. 293-315.
Von Eckardt, B. (1994), *Folk psychology*, in S. Guttenplan (a cura di), *A Companion to the Philosophy of Mind*, Blackwell, Oxford.
Watson, J. (1930), *Behaviorism*, University of Chicago Press, Chicago [trad. it. *Il comportamentismo*, Giunti-Barbera, Firenze 1984].
Waskan, J., Bechtel, W. (1997), *Directions in connectionist research: tractable computations without syntactically structured representations*, in «Metaphilosophy», 28, pp. 31-62.
Wason, P. (1966), *Reasoning*, in B. Foss (a cura di), *New Horizons in Psychology*, Penguin, Harmondsworth [trad. it. *I nuovi orizzonti della psicologia*, Boringhieri, Torino 1968].
Wellman, H.M., Cross, D., Watson, J. (2001), *Meta-analysis of theory-of-mind development: the truth about false belief*, in «Child Development», 72, pp. 655-684.
Wilson, R. (1999), *Philosophy*, in R. Wilson e F. Keil (a cura di), *The MIT Encyclopedia of the Cognitive Sciences*, MIT Press, Cambridge (MA).
Wilson, T. (1985), *Strangers to ourselves: the origins and accuracy of beliefs about one's own mental states*, in J. Harvey e G. Weary (a cura di), *Attribution in Contemporary Psychology*, Academic Press, New York.
Wilson, T. (2002), *Strangers to Ourselves. Discovering the Adaptive Unconscious*, Harvard University Press, Harvard.
Wimmer H., Perner, J. (1983), *Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception*, in «Cognition», 13, pp. 103-128.

L'autore

Massimo Marraffa (Roma, 1961) svolge attività di ricerca presso il Dipartimento di Filosofia dell'Università Roma Tre.

È autore di vari articoli di filosofia della mente e della scienza cognitiva. Per la rivista «Sistemi Intelligenti» ha curato un numero monografico dedicato alla psicologia dell'autoinganno e uno dedicato alla psicologia ingenua.

Ha contribuito ai volumi collettanei *Fondamenti di cognitivismo clinico*, a cura di C. Castelfranchi, F. Mancini e M. Miceli (Bollati Boringhieri, Torino 2001); *Storia della filosofia analitica*, a cura di F. D'Agostini e N. Vassallo (Einaudi, Torino 2002); *Normatività, fatti, valori*, a cura di R. Egidi, M. Dell'Utri e M. De Caro (Quodlibet, Macerata 2003).

È autore del saggio *Scienza cognitiva. Un'introduzione filosofica* (Cleup, Padova 2002²). Un altro volume, dal titolo *Oltre il riduzionismo. Un modello pluralista delle relazioni tra psicologia e neuroscienza*, è in preparazione presso la Bollati Boringhieri.

Indice

Premessa VII

1. Il progetto computazionale in psicologia 3
   1. La psicologia del senso comune *p. 4*
   2. Dentro la scatola nera: le vicissitudini dell'informazione *8*
   3. Il funzionalismo computazionale e rappresentazionale *17*
   4. Dire di più di quello che possiamo sapere *38*

2. La spiegazione psicologica e la struttura della mente 42
   1. Dalle leggi ai meccanismi *p. 42*
   2. La modularità della mente *55*
   3. Il problema della cognizione centrale *71*

3. Psicologia ingenua e razionalità naturale 76
   1. La psicologia della psicologia ingenua *p. 76*
   2. Razionalità normativa e razionalità naturale *95*

4. Psicologia e neuroscienza 109

1. L'autonomia della psicologia *p. 110*
2. Computazioni cerebriformi *114*
3. La dialettica della scomposizione e della localizzazione *128*

Cos'altro leggere 139

Bibliografia 145

L'autore 159