

Tutti i diritti riservati – All rights reserved
© 2006 Stamen – Edizioni scientifiche
Via A. Poliziano, 24 – 00184 Roma
e-mail: redazione@stamen.net
www.stamen.net

IDENTITÀ E
RAPPRESENTAZIONE
Scienza cognitiva e teorie della mente

Testi di
R.M. Gordon, S. Horst, D. Pitt
J.J.C. Smart, P. Thagard

Introduzione e cura di
Gloria Galloni

Con un saggio di
Enrica Vizzinisi



Prima edizione: giugno 2006

© 2006 Stamen – Edizioni scientifiche
Via A. Poliziano, 24 – 00184 Roma
e-mail: redazione@stamen.net
www.stamen.net

Traduzione a cura della
Scuola Superiore per Mediatori Linguistici (SSML)
«Università Gregorio VII»
Progetto grafico di copertina: © Stamen – Edizioni scientifiche
In copertina: Paul Klee, *Senecio* (1922), Kunstmuseum, Basilea

ISBN 88-901301-1-3

Indice

- 5 GLORIA GALLONI, *Introduzione*
- 23 PAUL THAGARD, *Introduzione alla scienza cognitiva*
- 41 J.C.C. SMART, *La teoria dell'identità della mente*
- 71 DAVID PITT, *La rappresentazione mentale*
- 103 STEVEN HORST, *La teoria computazionale della mente*
- 130 ROBERT M. GORDON, *Folk Psychology come simulazione
mentale*
- 141 ENRICA VIZZINISI, *Cognizione e significato*

Quando si parla di scienza cognitiva s'intende quella confluenza di ricerche multidisciplinari aventi ad oggetto i processi cognitivi. Nella scienza cognitiva s'incontrano studiosi di intelligenza artificiale, neuroscienze, filosofia, psicologia, linguistica ed antropologia, per citare solo gli ambiti disciplinari che presentano i maggiori apporti. I settori sono dunque molti ed eterogenei, tanto che spesso si è ritenuto opportuno parlare di «scienze cognitive», al plurale. Probabilmente a ragione, vista la problematicità epistemologica che l'unificazione sotto una stessa etichetta può comportare. Da un lato, infatti, i vari studiosi, pur provenienti da settori differenti, hanno in comune l'interesse per l'analisi della cognizione intelligente; dall'altro, a seconda della provenienza culturale e scientifica di ogni studioso, tale analisi sarà affrontata con metodi talvolta molto differenti. Questo pluralismo metodologico viene salutato da alcuni studiosi¹ come un fattore positivo di scambio, dialogo ed incontro, ma nasconde diverse insidie. Insidie che stanno portando, nell'ultimo decennio, ad un gran fervore ed all'annuncio di vere e proprie crisi nel paradigma kuhnianamente inteso. Proviamo a calarci nel contesto storico-teorico ed a comprendere ciò che accade.

¹ McCauley R. N., Bechtel W., *Explanatory pluralism and the heuristic identity theory*, in «Theory and Psychology», vol. 11, 2001, pp. 736-760.

1. *Le origini*

La scienza cognitiva nasce dall'incontro di alcune aree di ricerca che si trovano in accordo sull'uso di alcuni strumenti teorici nel discutere d'intelligenza e di processo cognitivo. Fondamentalmente furono tre i settori che inizialmente confluirono nel creare le condizioni storico-teoriche affinché la scienza cognitiva potesse nascere. Tali settori furono l'intelligenza artificiale, il cognitivismo e la linguistica generativa.

Com'è noto, la nascita dell'intelligenza artificiale viene datata convenzionalmente nel 1956, in occasione di un seminario che ebbe luogo a Dartmouth «sulla base della congettura che ogni aspetto dell'apprendimento o ogni altro carattere dell'intelligenza possa essere in linea di principio descritto in modo sufficientemente preciso, in modo che una macchina possa simularlo»² e con lo scopo di guardare «alla costruzione di macchine intelligenti [...] come alla migliore scommessa, sia nella prospettiva della costruzione di una intelligenza artificiale, sia in quella della comprensione dell'intelligenza naturale»³. L'idea di base di questo tipo di approccio è dunque una stretta analogia tra il funzionamento della mente umana ed il modo in cui un computer processa l'informazione, un'analogia che riveste tutt'oggi una grande importanza nelle scienze della mente, ma che ha condotto anche ad un recupero di terreno da parte dell'intuizione dualista. Come argomenta Domenico Parisi, infatti, «con l'idea della mente "computazionale" i cognitivisti hanno potuto incassare il vantaggio di considerare la mente come una macchina e quindi

² Documento preparatorio per il seminario di Dartmouth, reperibile al sito <http://www-formal.stanford.edu/jmc/>; trad. it. di G.I. Giannoli, in «Intelligenza artificiale e filosofia», in G. Fornero, S. Tassinari (a cura di), *Le Filosofie del Novecento*, Paravia, Bruno Mondadori Editore, Milano 2002, pp. 1443-1463.

³ McCarthy J., Hayes P.J., *Some philosophical problems from the standpoint of Artificial Intelligence*, in «Machine Intelligence», 4, 1969; trad. it. di G.I. Giannoli, cf. nota 2.

come qualcosa di studiabile rigorosamente – contro Cartesio – senza però dover abbandonare il dualismo tra la mente e il corpo dato che l’analogia con quella macchina *sui generis* che è il computer permetteva loro di continuare a tenere la mente al riparo dalle scienze della natura – facendo così contento Cartesio»⁴.

Il cognitivismo è una corrente psicologica nata dalle ceneri (e come filiazione) del comportamentismo, il quale nello sforzo di rendere la psicologia una scienza oggettiva e sperimentale al pari della fisica (in un’epoca in cui non appariva ancora chiara la distinzione che le scienze biologiche e storiche devono possedere rispetto alle scienze fisiche, distinzione che ancora oggi fa fatica ad affermarsi⁵), rinnega che si possa discutere scientificamente dei processi e contenuti mentali e, scartando la mente in quanto «black box» di cui non si può far scienza, sceglie di soffermarsi sullo studio del comportamento manifesto. Senza entrare nei particolari storici che hanno portato alla nascita del cognitivismo, sottolineiamo che ad attenti studi non poté sfuggire il fatto che a stimoli simili spesso corrispondono risposte differenti: deve esserci una variabile che interviene tra *input* ed *output*, e deve trattarsi di un processo interno alla «black box». S’iniziano dunque a studiare con metodologie sperimentali (tra queste, una grande importanza riveste l’analisi dei tempi di reazione) queste «entità cognitive»: l’analisi dei costituenti e dei processi interni alla mente costituisce l’oggetto del cognitivismo, il quale viene sistematizzato teoricamente da Ulric Neisser solo nel 1967 con il testo *Cognitive Psychology*⁶.

La nascita di un nuovo modo di guardare al linguaggio, alle sue strutture di base ed al suo sviluppo ontogenetico, si ebbe con il notevole contributo teorico di Noam Chomsky e la

⁴ Parisi D., *Mente. I Nuovi Modelli della Vita Artificiale*, Il Mulino, Bologna 1999, p. 52.

⁵ Si veda in proposito Mayr E.W., *Biologia ed Evoluzione*, Bollati-Boringhieri, Milano 1982.

⁶ Neisser U., *Cognitive Psychology*, Englewood Cliffs, N.J., Prentice-Hall, 1967.

sua linguistica generativa. Egli, contrapponendosi energicamente già nel 1957 al comportamentista Skinner in una recensione al suo testo *Verbal Behaviour*, negli anni successivi sostenne l'esistenza di un vero e proprio «organo mentale» del linguaggio, sviluppatosi evolutivamente e presente nell'uomo in maniera innata⁷. Tale organo conterrebbe in sé la possibilità stessa della rapida acquisizione della lingua nei bambini grazie all'innata (anche se embrionale) conoscenza e capacità di estrazione di un insieme di regole grammaticali universali, ossia valide per ogni lingua esistente. Questo modello della mente fu presto esteso ad altri ambiti ad opera dei cognitivisti e dei filosofi, e rivestì un punto di riflessione fondamentale in quanto nuovo approccio a quello che, probabilmente a ragione, viene indicato come il processo cognitivo per eccellenza: il linguaggio.

Grazie al contributo di questi settori disciplinari, la scienza cognitiva poté muovere i primi passi. L'intelaiatura teorica sottesa agli studi della cognizione fu dunque caratterizzata fortemente dall'analogia mente-computer e dall'idea che la mente sia un elaboratore di informazioni, un processore computazionale. Il paradigma computazionale, unito ad una teoria rappresentazionale della mente, alla congettura della realizzabilità multipla ed all'ipotesi della modularità dei processi cognitivi (che, come vedremo, vede in Chomsky uno dei padri) furono le direttive teoriche di fondo – di cui ora parleremo – per gli scienziati cognitivi degli anni Settanta ed Ottanta del secolo scorso, fino a quando altri approcci ed altre scoperte non misero in crisi buona parte di tali fundamenta.

2. *La cornice filosofica della prima scienza cognitiva*

Prima di andare avanti con la storia è opportuno analizzare da vicino il paradigma della «prima» scienza cognitiva, per meglio comprenderne la crisi successiva. In ambito filosofico, fu il funzionalismo ad incarnare la concezione del rapporto

⁷ Chomsky N., *The Logical Structure of Linguistic Theory*, Chicago, University of Chicago Press, 1975.

mente-corpo più congruente alla scienza cognitiva, alla quale anzi offrì, nell'opera di Jerry A. Fodor, una esaustiva sistematizzazione dell'impianto teorico.

Per funzionalismo in filosofia s'intende l'idea, nata con Hilary Putnam negli anni Sessanta⁸, che gli stati mentali, pur essendo implementati cerebralmente, siano qualcosa di distinto dai corrispondenti stati cerebrali: secondo i funzionalisti, gli stati mentali sarebbero, infatti, proprietà funzionali degli stati cerebrali, con i quali sono in rapporto di uno-a-molti. Questo significa che lo stesso stato mentale «dolore» (per riprendere un esempio largamente utilizzato in quest'ambito) può trovare differenti implementazioni fisiche in specie differenti ed in individui diversi (e, Fodor aggiungerà⁹, anche nello stesso individuo in tempi differenti). Nasce così l'ipotesi della realizzabilità multipla, che si pone subito come ostacolo alle concezioni materialistiche del rapporto mente-corpo, sebbene esse vivano in quegli anni un momento di particolare vitalità.

Nella seconda metà degli anni Cinquanta, infatti, al funzionalismo portato avanti negli Stati Uniti da Carnap¹⁰ e Feigl¹¹ si associa il cosiddetto materialismo australiano di U. Place¹², Jack J.C. Smart¹³ e D. Armstrong¹⁴. Costoro promuovono la

⁸ Putnam, H., *Minds and Machines*, New York, New York University Press, 1960; in Putnam H., *Mente, Linguaggio e Realtà*, Adelphi, Milano 1987.

⁹ Fodor J.A., *Special sciences*, «Synthese», 28, 1974, pp. 77-115.

¹⁰ Carnap R., *An Excerpt from "Psychology in Physical Language"*, in Lycan W.G. (a cura di), *Mind and Cognition. A Reader*, Oxford, Blackwell, 1990, pp. 23-28.

¹¹ Feigl H., *Logical analysis and the psychophysical problem*, in «Philosophy of Science», 1, 1934, pp. 420-445.

¹² Place U.T., *Is consciousness a brain process?*, in «British Journal of Psychology», 47, 44-50, 1956.

¹³ Smart J.J.C., *Sensations and brain processes*, in «The Philosophical Review», n. 68, 1959, pp. 141-56.

¹⁴ Armstrong D., *A Materialist Theory of the Mind*, London, Routledge, 1968.

cosiddetta *teoria dell'identità tra mente e corpo*. Tale teoria postula una identità ontologica tra stati mentali e stati cerebrali, nonostante i due aspetti si presentino alla coscienza fenomenicamente distinti: i fulmini sono scariche elettriche, e tuttavia non ci si presentano come tali se non tramite un'analisi scientifica! Nata con Place (in particolare molto successo ebbe il suo saggio del 1956 *Is Consciousness a Brain Process?*), tale teoria venne resa nota negli U.S.A. tramite un ciclo di conferenze tenuto da Smart, il quale nel 1959 raccolse nel volume *Sensations and Brain Processes* le obiezioni – e le possibili risposte – maturate nelle varie università durante le conferenze. Da Place e Smart si discosta invece Armstrong, il quale raccoglie negli anni le suggestioni della teoria causale di Lewis¹⁵ e la terminologia di Feigl¹⁶, tanto che nel 1968 nel volume *A Materialist Theory of the Mind* modifica anche il nome della teoria in «materialismo dello stato centrale», per distinguersi dall'approccio materialistico dei comportamentisti che, come visto sopra, si curavano degli aspetti «periferici» del comportamento. Alla teoria dell'identità *di tipo* (come pure viene definita per distinguerla da differenti ipotesi d'identità di cui brevemente parleremo) si oppone dunque il funzionalismo, forte dell'argomento della realizzabilità multipla. Con tale argomento si vuole infatti sostenere l'irriducibilità dei predicati psicologici ai corrispondenti stati neurofisiologici¹⁷.

Nel pensiero di J.A. Fodor confluiscono in un quadro unitario le concezioni che formano la cornice teorica di quella

¹⁵ Lewis D.K., *An argument for the identity theory*, 1966; ora in Rosenthal D.M., *Materialism and the Mind-Body Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1971, pp. 162-171.

¹⁶ Si confronti Feigl H., *Mind-Body, not a Pseudo-Problem*, in Hook S. (a cura di), *Dimensions of Mind*, New York, New York University Press, 1960 pp. 24-36.

¹⁷ Per una trattazione della realizzabilità multipla, si veda Gallo-
ni G., «La realizzabilità multipla: un errore filosofico?», in *Dialegesthai. Rivista telematica di filosofia*, anno 7, 2005,
<http://mondodmani.org/dialegesthai/ggl01.htm>

che Parisi¹⁸ definisce come «scienza cognitiva computazionale» (o, secondo una differente terminologia, «classica»¹⁹): funzionalismo ed analogia mente-computer, teoria computazionale-rappresentazionale della mente, modularità.

Il nucleo del pensiero fodoriano è effettivamente costituito dall'unione di una teoria rappresentazionale della mente (da ora, TRM) con una teoria computazionale (TCM). Teniamo a mente, per ora, che per rappresentazione (della quale si possono dare varie definizioni e teorizzazioni²⁰) si deve intendere un'immagine che ci facciamo della realtà, che media dunque tra noi ed il mondo esterno e che viene ad essere manipolata nella formazione del pensiero. La TRM in Fodor assume una peculiare fisionomia: questa rappresentazione interna viene ad essere pensata come un simbolo-modello interno dell'ambiente esterno avente, al pari di un enunciato del linguaggio naturale, una struttura proposizionale propria di sintassi (la forma) e semantica (il contenuto). A suo parere dunque nella nostra mente è presente un vero e proprio «linguaggio del pensiero»: il *Mentalese*. Per quanto riguarda invece la TCM, essa si iscrive all'interno del paradigma computazionale che, a partire dagli anni Cinquanta e dagli studi relativi all'intelligenza artificiale, come abbiamo visto, s'impone con il suo linguaggio e con le sue metafore anche nell'ambito dello studio della mente. In particolare, ribadiamo, la TCM può essere definita come la visione secondo cui nell'uomo l'atto di pensare equivale all'eseguire una serie di calcoli: l'uomo come elaboratore di informazioni, la mente come un computer. Unire TRM e TCM significa postulare che la forma (e cioè il livello sintattico) della rappresentazione mentale possa essere inserita in un processo di calcolo. Ciò vuol dire che, se il ragionamento è costituito da molteplici funzioni

¹⁸ Parisi, D. (2002), *Sulla scienza cognitiva*, in «Sistemi Intelligenti», anno XIV, n. 1.

¹⁹ SI veda Marras M., *Scienza Cognitiva. Un'Introduzione Filosofica*, CLEUP, Padova 2002.

²⁰ In quest'ambito, ci correrà in soccorso il saggio di D. Pitt contenuto in questo volume.

computabili, i costituenti di tali funzioni sono le proprietà formali dei contenuti del pensiero. Di qui, la possibilità d'implementare la capacità logica su un dispositivo meccanico quale la Macchina di Turing (Universal Turing Machine).

Fodor ritiene fondamentale, in sintonia con la psicologia del senso comune²¹, il ruolo causale degli stati mentali nel muovere all'azione. Il problema diviene allora conciliare tale mentalismo con una naturalizzazione del mentale, necessaria per evitare d'incorrere nel dualismo di stampo cartesiano (dualismo delle sostanze). Il ricorso al funzionalismo (studiare la mente ed i processi cognitivi nelle loro funzioni, senza che questo debba essere accompagnato da un'analisi del substrato su cui tali funzioni sono implementate) permette a Fodor di uscire dall'*empasse*, abbracciando la tesi della realizzabilità multipla del mentale e giungendo a postulare una forma di fisicalismo di tipo non riduzionista: la teoria dell'identità delle occorrenze. Con tale teoria si suppone che uno stesso *tipo* mentale quale il «dolore» venga di volta in volta (nelle sue varie *occorrenze*) implementato in substrati differenti. Un fisicalismo, dunque, ma non lontano da un dualismo delle proprietà (secondo il quale, pur non essendoci due tipi ontologici di realtà, tuttavia la realtà mentale non è riducibile alla realtà fisica possedendo proprietà aggiuntive rispetto ad essa).

Ma nella prassi, in che modo il linguaggio formale (la parte sintattica delle rappresentazioni) riesce a causare output comportamentali? La sintassi diviene causa di output con la computazione stessa (per lo stesso processo che conduce il computer che ho davanti agli occhi ad effettuare determinati passaggi essendo fornito di un programma che funziona in base a regole «sintattiche» e di determinati input ambientali – il mio premere certi tasti) preservando nei vari passaggi schemi

²¹ Un intero saggio della presente antologia (quello di Robert M. Gordon) è dedicato alla psicologia del senso comune. Per ora, ci basti definirla sommariamente come lo studio del modo in cui le persone si spiegano e comprendono gli atteggiamenti mentali altrui (le credenze, i bisogni, i desideri e così via).

di argomentazione valida che operano in modo tale da mantenere quella logicità e razionalità contenutistica che è propria dell'azione indirizzata ad uno scopo. Ed in che modo gli input ambientali confluiscono nel linguaggio del pensiero, subiscono manipolazioni e generano credenze e comportamenti? Nel famoso saggio *The Modularity of Mind* del 1983²², Fodor risponde in maniera organica a tali interrogativi parlando di un'ipotesi di architettura cognitiva che ha goduto, e tutt'ora gode, di una grande fortuna.

La modularità della mente viene affermata per indicare la possibilità di scomporre funzionalmente la capacità cognitiva in sistemi differenti, per cui ogni modulo viene ad indicare un certo meccanismo con una sua specifica e ben individuata funzione di elaborazione cognitiva. Nello specifico, i moduli fodoriani sono moduli chomskiano-computazionali: chomskiani in quanto basi di conoscenze innate e specifiche per dominio (che organizzano ed analizzano solo un certo tipo di input e sono in grado di acquisirne informazioni su cui svilupparsi a seguito di una fissazione di parametri, come appunto nella teoria del linguaggio di Noam Chomsky che abbiamo brevemente analizzato); computazionali in quanto elaborano in modo algoritmico un tipo di input specifico per dominio. La specificità per dominio (secondo cui i moduli analizzano solo un ristretto e ben definito tipo di input) e l'incapsulamento informazionale (secondo cui le informazioni situate all'esterno di un modulo non sono ad esso accessibili) sono le due proprietà fondamentali affinché si possa parlare di modulo. Un aspetto importante e distintivo dell'ipotesi fodoriana è che, a suo parere, la modularità è propria unicamente dei sistemi periferici, cioè dei sistemi di input, i quali si trovano a diretto contatto con i trasduttori dell'informazione ambientale (i sensi). I sistemi centrali, invece, nei quali si forma il «pensiero» propriamente detto, possiedono caratteristiche marcatamente differenti rispetto ai sistemi modulari e devono, per la natura delle elaborazioni cui devono dare luo-

²² Fodor J.A., *The Modularity of Mind. An Essays on Faculty Psychology*, Cambridge, Mass, The MIT Press, 1983.

go (per esempio, nel problem solving) poter essere organizzati olisticamente. Non è possibile qui entrare maggiormente nel merito di una trattazione esaustiva della modularità. Ci limitiamo ad osservare come essa sia del tutto coerente con il paradigma computazionale ed il funzionalismo e sembri anzi legittimarne lo statuto epistemologico. Inoltre, chiunque abbia un minimo di dimestichezza con la neuropsicologia può facilmente constatare quanto la teoria della modularità offra ad essa un buon modello filosofico in accordo con gli studi sui deficit cognitivi: la neuropsicologia correla infatti la lesione in un punto specifico del cervello ad un'altrettanto specifica perdita di una determinata funzione cognitiva. Si presuppone dunque che vi siano aree cerebrali particolari per ciascuna attività funzionale, ovvero che ci sia una generale modularità: una suddivisione del cervello in tante parti funzionalmente diversificate e selettivamente danneggiabili. Si apre così lo scenario ad un dialogo interessante tra la filosofia e le neuroscienze.

3. *La crisi dei fondamenti: una seconda scienza cognitiva?*

A partire dalla fine degli anni Ottanta venne sempre più in luce una certa insofferenza per il divario tra gli studi di scienza cognitiva e le scoperte delle neuroscienze, le cui nuove tecniche di *imaging* cerebrale iniziavano a permettere studi del tutto nuovi dei processi cognitivi, di un rilievo prima solo fantasticato. Nelle parole di Pavlov, ottant'anni fa: «Se fosse possibile vedere attraverso la scatola cranica e se la zona maggiormente eccitata fosse luminosa, si potrebbe seguire, in un uomo intento a pensare, lo spostamento incessante di questo punto luminoso, in un continuo cambiamento di forma e dimensione, e circondato da una zona d'ombra più o meno fitta che occuperebbe tutto il resto degli emisferi»²³. Questa speranza è ora una realtà dall'immenso valore euristico.

²³ Pavlov I.P., *Conditioned Reflexes*, London, Routledge and Kegan Paul, 1927.

Iniziarono inoltre ad essere prese in nuova considerazione le intuizioni di Donald O. Hebb²⁴, che già negli anni Cinquanta erano state incarnate nelle macchine costruite da Oliver Selfridge e Frank Rosenblatt: il *Pandemonium* ed il *Perceptron*. Dopo vent'anni da questi primi approcci connessionistici, nacquero le reti neurali artificiali, per costruire le quali si basa sul modello biologico cerebrale. Esse consistono in una serie di unità, paragonabili funzionalmente ognuna ad un neurone, collegate tra di loro in modo che possano passarsi dei «pesi» corrispondenti al livello di attivazione neurale. Tali unità sono strutturate, nelle reti più semplici, in tre livelli: unità di input, unità nascoste ed unità di output. Il collegamento, i pesi e le risultanti di tutti questi rapporti generano una certa risposta da parte della macchina. Non è questa la sede per entrare nei dettagli di tale approccio all'intelligenza artificiale, un approccio molto potente (basti pensare che una macchina dotata di reti neurali è capace di apprendere) che ha generato l'ipotesi che si fosse di fronte ad un cambiamento nel paradigma computazionale²⁵. Quel che dal nostro personale punto di vista appare interessante è che in questo modo si è cercato di porre rimedio, pur sempre all'interno di un quadro teorico *computazionale* (il pensiero è sempre immaginato come manipolazione di informazioni, seppure in un formato sub-simbolico), al *gap* tra le ricerche della psicologia cognitiva e dell'intelligenza artificiale di cui finora abbiamo discusso e le ricerche neurofisiologiche sul sistema nervoso centrale.

Anche la filosofia, d'altra parte, inizia a riflettere sulle neuroscienze cognitive e sul ruolo della localizzazione cerebrale e dei deficit cognitivi per lo studio della mente che le

²⁴ Si pensi in particolare al concetto di «apprendimento dobbiamo», secondo cui l'apprendimento viene rinforzato dal ripetersi dell'attivazione sincrona dei corrispondenti gruppi di neuroni, ovvero di quelle che lui definisce «assemblee cellulari»; in Hebb D.O., *The Organization of Behavior*, New York, John Wiley Inc., 1949.

²⁵ Si veda P. Smolensky (1988), *Il Connessionismo tra Simboli e Neuroni*, Marietti, Genova 1992.

nuove tecnologie permettono in modo sempre più particolareggiato²⁶. Abbiamo parlato del ruolo della modularità come modello teorico per la localizzazione delle funzioni in neuropsicologia. Essa rende ragione della selettività dei deficit cognitivi, e viene dimostrata tramite il metodo sottrattivo e la doppia dissociazione abbinate ad una registrazione effettuata tramite FMRI²⁷. In questo contesto la modularità di stampo fodoriano cede il passo ad una modularità ‘allargata’ non solo ai trasduttori di input ma a tutta la cognizione, non senza problematicità: se anche i sistemi cognitivi ‘superiori’ sono modulari²⁸, e come tali incapsulati, in che modo i moduli riescono a correlarsi dando luogo a quei fenomeni di contestualizzazione che innegabilmente sono alla base della generazione, ad esempio, di credenze da parte del soggetto? Su questo aspetto sono centrate gran parte delle riflessioni attuali inerenti la modularità.

L’urgenza di allontanarsi da questo *solipsismo metodologico*²⁹, ovvero dall’idea che al fine di comprendere la cognizione si debbano studiare soltanto i processi di elaborazione dell’informazione aventi luogo nella ‘testa’ dell’individuo, si unì ben presto alla necessità di riavvicinare studio della mente e studio del corpo, in particolare del sistema nervoso cen-

²⁶ Per una dettagliata storia dei modelli di funzionamento della mente dall’antichità ad oggi, con particolare riferimento alla neuropsicologia, si veda Morabito C., *La Mente nel Cervello*, Laterza, Roma-Bari 2004.

²⁷ Si veda in proposito l’esempio di doppia dissociazione tra pazienti prosopoagnosici e Capgras riportato da Marraffa e Meini nel testo *La Mente Sociale. Le Basi Cognitive della Comunicazione*, Laterza, Roma-Bari 2005.

²⁸ Su questo, si pensi all’ipotesi di modularità massiva di Dan Sperber, *The modularity of thought and the epidemiology of representations*, in L.A. Hirschfeld, S. A. Gelman (a cura di), *Mapping the Mind: Domain Specificity in Cognition and Culture*, Cambridge, Cambridge University Press, 1994, pp. 39-67.

²⁹ Si veda in proposito Dreyfus H.L., (1972), *Quel che i Calcolatori non Possono Fare*, Armando, Roma 1988.

trale. Questo fu reso possibile da un insieme di fattori convergenti.

In primo luogo, già all'interno della psicologia cognitivista, una delle correnti portanti della scienza cognitiva classica, si iniziò ad avvertire un certo disagio per quell'uso esasperato di *modelli* staccati da un'effettiva considerazione della fisiologia ed anatomia da una parte e del contesto adattativo nel quale la mente si è evoluta dall'altro. L'approccio ecologico di Gibson³⁰ è senz'altro un sintomo di questo malessere e della volontà di reinserire la mente all'interno del corpo e dell'ambiente, con il quale, secondo Gibson, abbiamo un rapporto diretto, non mediato da rappresentazioni e processi intermedi di calcolo: cogliamo opportunità d'azione nell'ambiente (*affordances*), e le cogliamo in modo originario.

Un ruolo di primo piano in questo processo lo dobbiamo riconoscere inoltre all'enorme sviluppo delle già citate tecniche di *brain imaging*, il cui potere euristico fu da subito avvertito come dirompente. Le tecniche di *neuroimaging* hanno comportato una vera e propria rivoluzione negli studi cognitivi, perché permettono la visualizzazione *in vivo* dell'attività cerebrale.

Nel campo squisitamente filosofico, sempre più spesso si iniziò a rivalutare il pensiero di quei filosofi che, sin dall'inizio del secolo scorso, posero l'accento sull'uomo come essere-nel-mondo (in particolare, Heidegger³¹) e sul ruolo della corporeità propria (Husserl³²) e del vissuto fenomenologico-

³⁰ Gibson J., *The Ecological Approach to Visual Perception*, London, Erlbaum, 1979.

³¹ Heidegger M., *Sein und Zeit*, Tübingen, Max Niemeyer, 1927; trad. it. *Essere e Tempo*, Longanesi, Milano 1968.

³² Husserl E., *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*, Erstes Buch: Allgemeine Einführung in die reine Phänomenologie, 1913, trad. it. *Idee per una Fenomenologia Pura e per una Filosofia Fenomenologica*, Einaudi, Torino 2002.

esperienziale nella cognizione (Merleau-Ponty³³). Lo stesso approccio computazionale nell'intelligenza artificiale virò verso una modellistica che ricalcasse l'effettivo riferimento biologico, e dunque si assistette, come sopra annunciato, alla nascita delle reti neurali artificiali.

Persino la robotica è stata influenzata dal convergere di tali e tanti fattori, e con la nuova robotica (si pensi a Brooks³⁴ e Clark³⁵) l'ambiente è considerato parte del soggetto agente, il quale si poggia su di esso come *impalcatura cognitiva* attraverso il proprio corpo. I nuovi *mobot* (robot semoventi ed autonomi) sono forniti di un apparato cognitivo privo di una netta distinzione tra percezione, cognizione ed azione, ma il cui cervello – simulato da una rete di tipo connessionistico – è formato da sottosistemi che sono percorsi completi dall'input all'azione. Il repertorio comportamentale, dunque, dell'agente cognitivo è pronto ad attivarsi su «richiamo» degli input ambientali (come nelle *affordances* gibsoniane).

Soprattutto nell'ultimo decennio, importanti ricerche neurofisiologiche portate avanti dal gruppo di Parma (Rizzolatti, Gallese, Fogassi e gli altri) hanno condotto alla scoperta di neuroni multimodali che aprono la strada a nuove teorizzazioni concernenti il «formato» della categorizzazione ed il ruolo sempre più rilevante, per la semantica e la concettualizzazione, della sensorialità e del corpo in movimento³⁶. Tutto questo convergere sull'importanza dello studio del corpo, del cervello, dell'azione e dell'ambiente fa sì che si sia parlato di

³³ Merleau-Ponty M., *Phénoménologie de la Perception*, Paris, Gallimard, 1945; trad. it. *Fenomenologia della Percezione*, Il Saggiatore, Milano 1965.

³⁴ Si veda Steels L., Brooks R.A. (a cura di), *The Artificial Life Route to Artificial Intelligence: Building Embodied Situated Agents*, Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1995.

³⁵ Clark A., *Being There: Putting Brain, Body, and World Together Again*, MIT Press, 1997; trad. it. *Dare Corpo alla Mente*, McGraw-Hill Italia, Milano 1999.

³⁶ Rizzolatti G., Sinigaglia C., *So quel che Fai. Il Cervello che Agisce e i Neuroni Specchio*, Raffaello Cortina Editore, Milano 2006.

una seconda scienza cognitiva, una scienza cognitiva post-classica. In essa, il riferimento teorico di base e la terminologia più diffusa non saranno più quelli dell'intelligenza artificiale ma quelli delle neuroscienze: «Per la scienza cognitiva computazionale la mente è come il software di un computer, è un sistema computazionale, cioè un sistema di simboli e di regole per manipolare questi simboli. Di conseguenza, come il software del computer può e deve essere studiato indipendentemente dall' hardware in cui il software "gira", così la mente può e deve essere studiata ignorando la macchina fisica che le fa da supporto, cioè il cervello e più generalmente il corpo [...] Per la scienza cognitiva neurale la mente non ha nulla a che fare con il software di un computer, non è un sistema computazionale e per capire la mente bisogna partire dal cervello e dal corpo e dalle scienze che studiano il cervello e il corpo»³⁷.

4. *L'importanza della scienza cognitiva classica*

Sulla strada che ha condotto ad una tale rivalutazione del corpo e dell'ambiente per lo studio della mente, la scienza cognitiva classica è stata una tappa fondamentale ed irrinunciabile. È la scienza cognitiva classica, infatti, che per prima ha permesso una comunicazione così potente tra ambiti disciplinari diversi. È grazie ad uno studio trasversale della mente che si è potuta avvertire tutta l'utilità e la valenza euristica dell'incontro tra studiosi di formazione diversa su un terreno comune, con obiettivi comuni. Ed è sempre all'interno di questo approccio alla mente che si sono potuti avvertire quei limiti che hanno portato all'apertura di uno studio della mente in verticale, verso il cervello, ed in orizzontale, verso l'ambiente³⁸.

In questa antologia, ricca di contributi importanti nel panorama della filosofia della mente all'interno della scienza

³⁷ Parisi, op. cit. in nota 18, pp. 112-113.

³⁸ La distinzione «verticale-orizzontale» è presentata da Marraffa; si veda il testo citato nella nota 19.

cognitiva, scelti fra alcuni dei testi più significativi pubblicati negli ultimissimi anni (2004; 2005) nella *Stanford Encyclopedia of Philosophy*, si è tentato di eguagliare il fare delle api: all'interno di quell'orizzonte e dei suoi limiti, che ci hanno permesso di estendere lo sguardo verso nuovi e trascurati luoghi, vorremmo prestare attenzione a varie tematiche, traendone un'immagine del mondo (o meglio, della mente), un significato, un'idea da rivisitare, da elaborare e con la quale creare un nuovo nettare, un nettare dal sapore interdisciplinare, con una forte connotazione corporea, oltre che cognitiva. Perché non siamo menti disincarnate, e dobbiamo sempre tenere presente il nostro carattere «environmentally embedded, corporeally embodied and neurally “embrained”»³⁹.

Ma dalle radici bisogna partire, e dunque noi partiamo dalle idee delle grandi personalità filosofiche che nella scienza cognitiva classica hanno creduto, che di essa hanno vissuto. Il nostro sguardo è verso il futuro, ma consapevoli dell'importanza del passato: delle teorie che, dalla metà del secolo scorso fino agli Ottanta, hanno prevalso nello studio della mente. Saremo allora curiosi di immergerci – con Paul Thagard – nelle questioni metodologiche e nelle sottili sfumature teoriche che convergono nello studio interdisciplinare della cognizione. Jack J.C. Smart ci parlerà da interprete diretto della teoria dell'identità di tipo (importante ed interessante la sua analisi del rapporto tra di essa ed il funzionalismo), a favore ed in difesa della quale si è diretta la sua opera di diffusione negli *States*: una concezione del rapporto mente-corpo che ancor oggi altri importanti studiosi di filosofia delle neuroscienze prendono seriamente in considerazione come modello euristico con il quale studiare la mente-cervello⁴⁰. La

³⁹ Van Gelder T., *Dynamic approaches to cognition*, <http://www.arts.unimelb.edu.au/~tgelder/papers/MITDyn.pdf>, in Wilson R., Keil F. (a cura di), *The MIT Encyclopedia of Cognitive Sciences*, <http://mitpress.mit.edu./mitecs>, Cambridge MA, MIT Press, 2003, pp. 244-246.

⁴⁰ Bechtel W., *Heuristic identity theory (or back to the future): the mind-body problem against the background of research strategies in*

rappresentazione mentale è al centro del saggio di David Pitt, che analizzerà per noi le definizioni che se ne possono dare in ancoraggio alle diverse teorie della mente, soffermandosi inoltre sulle questioni che da tali divergenze teoriche possono sorgere. Oltre la teoria rappresentazionale, l'altro polo fondamentale che fa da sfondo alla scienza cognitiva classica è la teoria computazionale esaminata qui da Steven Horst, con importanti riferimenti alla storia della logica formale e matematica e la sua applicazione nel campo filosofico-psicologico e dell'intelligenza artificiale. Inoltre, Horst apre alle questioni critiche che tale interpretazione computazionale del processo cognitivo porta con sé, e al modo in cui questi problemi hanno condotto a nuovi approcci, da quelli connessionistici alla nuova robotica. Robert M. Gordon incentra il suo saggio sulla *folk psychology*, ovvero su quell'insieme di conoscenze di «psicologia quotidiana» o «ingenua» grazie alle quali comprendiamo il significato delle azioni altrui, sulle quali dunque si fonda la nostra dimensione di esseri sociali. In particolare, Gordon fa un salto qualitativo interpretando la possibilità di attribuire ad altri credenze, desideri ed intenzioni in quanto basata sulla possibilità di simulare internamente tali processi psicologici; ciò, lo sottolineiamo volentieri, è in pieno accordo con le scoperte del gruppo di Parma (Gordon, infatti, cita Gallese e Goldman⁴¹). Sulla scoperta dei *neuroni specchio*, che si attivano sia quando si compie un'azione che quando si osservano altri compierla – un atto, dunque, di simulazione – si poggiano le nuove teorizzazioni concernenti l'imitazione, l'empatia e la comprensione dei comportamenti altrui⁴².

cognitive neuroscience, in Hahn M., Stoness S.C. (a cura di), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society*, Mahwah, Nj, Lawrence Erlbaum Associates, 1999.

⁴¹ Gallese V., Goldman A., *Mirror neurons and the simulation theory of mind-reading*, in «Trends in Cognitive Sciences», vol. 2, n. 12, 1998, pp. 493-501.

⁴² Si confronti il testo di Rizzolatti e Sinigaglia citato in nota 36; si confronti inoltre l'articolo di Walter H., Adenzato M., Ciaramidaro A., Enrici I., Pia L., Bara B., *Understanding intentions in social*

Il percorso antologico che abbiamo scelto ci aiuterà a capire – questa la nostra speranza – in che modo, dall'interno stesso della scienza cognitiva classica, si è giunti a comprendere i limiti di un approccio centrato sulla metafora mente-computer e la necessità di modificare questo paradigma inglobando altri approcci allo studio della mente: la nuova robotica, le neuroscienze, l'ipotesi di *embodied cognition*, le teorie dei sistemi dinamici. Infatti, come sottolineato in un recente articolo pubblicato sulla rivista *Brain and Cognition*, «la seconda generazione della scienza cognitiva differisce dalla prima non solo nel suo ripudio del funzionalismo computazionale ma anche nella nuova concezione del soggetto che le è proprio, la cognizione *umana*. Invece di astratti processi mentali, descrivibili nei termini della logica formale, i processi cognitivi sono considerati alla luce dei loro effettivi legami all'*azione del corpo* ed all'*esperienza senso-motoria*»⁴³.

Nell'appendice abbiamo dunque scelto di presentare la fisionomia e la potenza euristica del nuovo paradigma attraverso un saggio di Enrica Vizzinisi, la quale, forte dell'approccio classico, conduce il proprio sguardo verso il nuovo orizzonte e verso tutto ciò che di positivo il fiorentino incontro tra le discipline costituenti la nuova scienza cognitiva può portare.

Gloria Galloni
Giugno 2006

interaction: the role of the anterior paracingulate cortex, in «Journal of Cognitive Neuroscience», 16:10, pp. 1854-1863, 2004.

⁴³ F. Garbarini, M. Adenzato, "At the root of embodied cognition: cognitive science meets neurophysiology", *Brain and Cognition*, 56, pp. 100–106, 2004; trad. e corsivo della curatrice.

INTRODUZIONE ALLA SCIENZA COGNITIVA⁴⁴

Paul Thagard

Possiamo definire la «scienza cognitiva» come uno studio interdisciplinare delle facoltà intellettive in cui si fondono filosofia, psicologia, intelligenza artificiale, neuroscienze, linguistica e antropologia. Le sue origini risalgono alla metà degli anni '50, quando alcuni ricercatori provenienti da diversi settori scientifici cominciarono ad interessarsi allo sviluppo di teorie sulla mente basate su complesse rappresentazioni e procedure computazionali. Tuttavia, solo verso la metà degli anni '70 è stata creata la «Cognitive Science Society», seguita poi dalla prima rivista specializzata, «Cognitive Science». Da allora, negli Stati Uniti ed in Europa, più di sessanta università hanno attivato programmi di scienza cognitiva e molte altre hanno organizzato corsi di insegnamento relativi a questa disciplina.

1. *Cenni storici*

I primi tentativi di comprendere la mente e i suoi meccanismi risalgono all'antica Grecia, quando filosofi come Platone ed Aristotele cercarono di spiegare la natura del sapere umano. Lo studio della mente ha rappresentato una prerogativa della filosofia sino al XIX secolo, momento in cui ha cominciato a svilupparsi la psicologia sperimentale. Wilhelm

⁴⁴ Paul Thagard (2004), *Cognitive Science*, in «Stanford Encyclopedia of Philosophy», <http://plato.stanford.edu/entries/cognitive-science/>

Wundt, insieme ad alcuni assistenti, elaborò una serie di metodologie finalizzate ad analizzare i processi cognitivi in modo sistematico. Dopo pochi decenni, la psicologia sperimentale venne criticata dal behaviorismo che negava l'esistenza della mente. Secondo i behavioristi come J.B. Watson, la psicologia dovrebbe limitarsi ad esaminare la relazione tra gli stimoli e le relative risposte comportamentali osservabili. In questo modo argomenti come le rappresentazioni mentali o la coscienza vennero banditi da qualsiasi rispettabile discussione scientifica. Specie nel Nord America, il behaviorismo ha dominato la scena lungo tutti gli anni '50, anche se a partire dal 1956 il panorama intellettuale ha cominciato ad evolversi significativamente. George Miller ha riassunto numerosi studi in cui si è dimostrato che la capacità del pensiero umano risulta essere limitata (per esempio a livello della memoria a breve termine, che può contenere al massimo sette dati). Egli sosteneva però che le limitazioni mnemoniche potessero essere superate grazie all'elaborazione dell'informazione in grandi blocchi, in rappresentazioni mentali che richiedono procedure mentali adeguate per codificare e decodificare le informazioni. In quel periodo la nascita dei primi calcolatori elettronici era appena avvenuta, ma pionieri come John McCarthy, Marvin Minsky, Allen Newell e Herbert stavano intanto individuando i campi dell'intelligenza artificiale. È importante sottolineare poi come Noam Chomsky rigettasse nettamente le teorie comportamentiste sul linguaggio inteso come un'abitudine appresa, e proponesse al contrario un'ipotesi secondo cui la struttura della competenza linguistica risulta basata sulle regole della grammatica generativa. I sei scienziati menzionati in questo paragrafo vanno considerati a tutti gli effetti i fondatori del cognitivismo.

2. I Metodi

Sebbene la scienza cognitiva presenti teorie alquanto omogenee, dobbiamo apprezzare la varietà di vedute e di metodi che i ricercatori dei diversi campi applicano allo studio della mente e dell'intelligenza. Nonostante alcuni psicologi

cognitivi spesso s'impegnino nella teorizzazione e nella modellazione computazionale, il loro metodo principale resta la sperimentazione diretta sull'uomo. Ai soggetti della sperimentazione, solitamente studenti universitari ritenuti idonei, viene monitorata, in laboratorio, qualsiasi variante di pensiero. Ad esempio, gli psicologi hanno esaminato sperimentalmente i diversi errori compiuti in un ragionamento deduttivo, i diversi modi in cui vengono formulati e applicati i concetti, la velocità dei pensieri con le immagini mentali, la risoluzione dei problemi con l'uso di analogie e molto altro. Le nostre conclusioni riguardo il funzionamento dell'attività mentale non possono basarsi solo sul «buon senso» o sull'introspezione, dal momento che questi ci offrono spesso un quadro ingannevole delle operazioni mentali, molte delle quali non sono accessibili consciamente. Gli esperimenti psicologici che toccano, prudentemente, le funzioni mentali da diversi punti di vista sono pertanto decisivi ai fini di rendere scientifica la scienza cognitiva.

Nonostante le teorie senza esperimento siano vuote, gli esperimenti senza teoria sono ciechi. Per far fronte agli interrogativi cruciali sulla natura della mente, gli esperimenti psicologici devono poter essere interpretati all'interno di una struttura teoretica che presuppone le rappresentazioni mentali e i procedimenti. Uno dei modi migliori per sviluppare strutture teoretiche è quello di formulare e verificare modelli computazionali affini alle funzioni mentali. Per completare gli esperimenti psicologici sul ragionamento deduttivo, sulla formazione di un concetto, sulle immagini mentali e sulle soluzioni ai problemi analogici, gli scienziati hanno sviluppato dei modelli computazionali che simulano alcuni aspetti del comportamento umano. Il creare, costruire e sperimentare con questi modelli computazionali costituisce il metodo centrale dell'Intelligenza Artificiale (IA), un ramo dell'informatica coinvolto nei sistemi d'intelligenza.

In linea teorica, nella scienza cognitiva i modelli computazionali e gli esperimenti psicologici procedono insieme, ma un importante studio dell'IA ha esaminato il potere dei diversi approcci alle rappresentazioni della conoscenza una volta iso-

lati dalla psicologia sperimentale. Mentre alcuni linguisti fanno esperimenti psicologici o sviluppano modelli computazionali, altri usano metodi alternativi. Per i linguisti della tradizione chomskiana, l'obiettivo teoretico principale è quello di identificare i principi grammaticali che forniscono le strutture-base del linguaggio umano. Il lavoro comincia col riconoscere la sottile differenza tra un'espressione grammaticale e una non grammaticale. In italiano ad esempio, la frase «Lei mangia una mela» e «Che dolce ti piace di più?» sono corrette, ma «Lei mela una mangia» e «Che dolce ti piacciono di più» sono errate grammaticalmente. Una grammatica italiana può spiegare perché le prime due frasi sono corrette e perché al contrario le ultime due non lo sono.

Come gli psicologi cognitivi, spesso i neuroscienziati effettuano esperimenti controllati, ma le loro osservazioni sono ben diverse se si considera il loro coinvolgimento con la natura del cervello. Con soggetti non umani, i ricercatori possono inserire elettrodi e registrare la scarica nel singolo neurone. Con gli esseri umani, nei confronti dei quali tale tecnica risulterebbe troppo invasiva, è stata introdotta negli ultimi anni la possibilità di usare un dispositivo magnetico o al positrone con lettura a scansione per verificare ciò che succede nelle varie zone del cervello mentre la persona elabora dei ragionamenti. Ad esempio, la scansione del cervello ha identificato la regione in cui avviene l'interpretazione della parola e in cui risiedono le immagini mentali. Ulteriori prove sulle funzioni del cervello sono raccolte dall'osservazione dei risultati di soggetti con danni cerebrali. Un ictus nella zona del cervello che controlla il linguaggio può produrre un deficit come quello di non riuscire più a costruire frasi corrette. Proprio come la psicologia cognitiva, la neuroscienza è spesso teoretica e sperimentale, e le teorie sono frequentemente aidate dallo sviluppo di modelli computazionali del comportamento di gruppi di neuroni.

L'antropologia cognitiva amplia lo studio del pensiero umano al fine di riuscire a comprendere come esso lavori nei differenti ambiti culturali. Risulta infatti ovvio come lo studio della mente non possa essere ristretto alla modalità del pen-

siero di chi parla inglese, e che debba ampliarsi a valutare e studiare le varie possibili differenze di pensiero delle varie culture. La scienza cognitiva sta divenendo sempre più consapevole della forte necessità di studiare i processi mentali in ambienti sia sociali che fisici particolari. Per ciò che riguarda gli antropologi culturali, uno dei metodi principali è quello dell'etnografia, che richiede un periodo di tempo di convivenza e interazione con i membri di una specifica cultura in maniera tale da comprendere pienamente i processi sociali e cognitivi della medesima. Gli antropologi cognitivi hanno studiato ad esempio le similitudini e le differenze tra le varie culture per quel che riguarda le parole che designano i colori.

A parte alcune eccezioni, i filosofi generalmente non si soffermano in sistematiche osservazioni empiriche o sulla costruzione di modelli di calcolo. Ma la filosofia resta importante per la scienza cognitiva in quanto tratta questioni fondamentali che sono alla base dell'approccio sperimentale e computazionale della mente. Questioni astratte quali la natura della rappresentazione e la computazione non debbono necessariamente essere indirizzati alla pratica psicologica o all'IA, ma ne derivano inevitabilmente quando i ricercatori scendono nelle profondità del loro studio.

La filosofia tratta interrogativi generali quali la relazione tra il corpo e la mente assieme ad interrogativi metodologici. Si interessa sia di questioni normative riguardo al modo in cui le persone dovrebbero pensare, sia di questioni descrittive su come concretamente pensano. Al di là dell'obiettivo teorico di comprendere il pensiero umano, la scienza cognitiva ha anche l'obiettivo pratico di migliorarlo, il che richiede una riflessione normativa su cosa desideriamo pensare di essere. La filosofia della mente non ha un metodo ben distinto, ma dovrebbe condividere i risultati empirici con i migliori lavori teoretici di altri campi.

Detto in modo più semplicistico, la scienza cognitiva è proprio *l'insieme di tutti gli ambiti* menzionati: psicologia, intelligenza artificiale, linguistica, neuroscienze, antropologia e filosofia. Un lavoro interdisciplinare diviene più interessante nel momento in cui la parte teoretica e quella sperimentale

convergono su una conclusione riguardante la natura della mente. Per esempio, la psicologia e l'intelligenza artificiale possono essere combinate attraverso modelli computazionali sul modo in cui le persone si comportano durante gli esperimenti. La maniera migliore per comprendere la complessità del pensiero umano è quella di utilizzare metodi multipli – ad esempio esperimenti psicologici e neurologici – e modelli computazionali. Teoreticamente l'approccio più promettente, per ciò che riguarda la comprensione della mente, si è avuto proprio in termini di rappresentazione e computazione.

3. Rappresentazione e computazione

Dall'ipotesi centrale della scienza cognitiva si evince che il pensiero può essere spiegato in termini di strutture mentali rappresentative e procedure computazionali che operano all'interno delle strutture stesse. Mentre ci sono molte discrepanze sulla natura delle rappresentazioni e delle computazioni che costituiscono il pensiero, l'ipotesi fondamentale è abbastanza generica da comprendere oggi diverse correnti di pensiero della scienza cognitiva, comprese le teorie connessionistiche che elaborano modelli del pensiero utilizzando le reti neurali artificiali.

Nel campo della scienza cognitiva, la maggior parte degli studi presuppone l'esistenza di rappresentazioni mentali analoghe alle strutture dei dati di un computer e procedure computazionali simili agli algoritmi computazionali. Secondo i teorici della scienza cognitiva, la mente contiene rappresentazioni mentali sotto forma di proposizioni logiche, regole, concetti, immagini e analogie e usa procedure mentali come la deduzione, la ricerca, il confronto, la successione e la memorizzazione. La nuova tendenza della scienza cognitiva allarga i confini dell'analogia dominante mente/computer al cervello.

Per ciò che concerne la rappresentazione e la computazione, i connessionisti hanno proposto idee innovative utilizzando i neuroni e le loro connessioni come modello per le strutture di dati e le loro attività elettrochimiche come fonte d'ispirazione per gli algoritmi. La scienza cognitiva si occupa

inoltre della triplice analogia mente-cervello-computer: ciascuno di questi può, singolarmente, suggerire nuove idee sullo studio degli altri. Non esiste in questo senso un singolo modello computazionale della mente, poiché diversi tipi di computer e di programmazione suggeriscono l'esistenza di differenti forme in cui la mente potrebbe operare. La maggior parte dei computer con cui oggi lavoriamo sono processori seriali che elaborano un comando alla volta, mentre il cervello e alcuni computer di nuova generazione sono definibili come processori paralleli, capaci di elaborare più dati contemporaneamente.

4. *Approcci teorici*

Lo schema che segue ci fornisce un sintesi delle attuali teorie sulla natura delle rappresentazioni e della computazione che mirano a spiegare il modo in cui lavora la mente.

4.1 *Logica formale*

La logica formale fornisce alcuni importanti strumenti per esaminare la natura delle rappresentazioni e della computazione. I calcoli proposizionali e predicativi servono per esprimere una serie di conoscenze complesse e molte inferenze possono essere interpretate in termini di deduzioni logiche con regole conclusive come *modus ponens*. Lo schema esplicativo per l'approccio logico è:

Obiettivo della spiegazione:

— Perché gli individui fanno delle deduzioni?

Concetto esplicativo:

— gli individui possiedono delle rappresentazioni mentali simili alle proposizioni predicative;

— gli individui possiedono procedure deduttive ed induttive che operano in queste proposizioni;

— le procedure deduttive ed induttive, applicate alle proposizioni, generano inferenze.

Non è accertato, tuttavia, che la logica fornisca le idee centrali riguardanti la rappresentazione e la computazione necessarie per la scienza cognitiva, poiché al fine di poter comprendere il pensiero umano risultano fondamentali anche dei metodi di computazione più efficienti e di natura psicologica.

4.2 Regole

La maggior parte della conoscenza umana viene descritta in termini di regole secondo la formula «se...allora...» e molte forme di pensiero come l'organizzazione possono modellarsi su sistemi basati sulle regole. Lo schema esplicativo in questo caso è:

Obiettivo della spiegazione:

— Perché gli individui possiedono un particolare tipo di comportamento intelligente?

Concetto esplicativo:

— gli individui possiedono delle regole mentali;
— gli individui possiedono le procedure per utilizzare queste regole, necessarie per individuare una serie di possibili soluzioni e procedure per la creazione di nuove regole;
— le procedure per l'utilizzo e la creazione di regole producono il comportamento.

I modelli computazionali basati sulle regole hanno fornito simulazioni dettagliate di un'ampia gamma di esperimenti psicologici, dalla risoluzione criptoaritmetica dei problemi all'acquisizione dell'uso del linguaggio. I sistemi basati sulle regole sono risultati importanti anche a livello pratico per

meglio comprendere l'apprendimento e per sviluppare modelli di macchine intelligenti.

4.3 Concetti

I concetti, che in parte corrispondono alle parole del linguaggio scritto e parlato, costituiscono importanti tipologie di rappresentazioni mentali. Esistono motivazioni computazionali e psicologiche che permettono di abbandonare la visione classica secondo la quale i concetti possiedono delle definizioni precise, anzi i concetti possono essere considerati come una serie di caratteristiche tipiche. L'applicazione dei concetti si riferisce alla creazione di un rapporto tra i concetti ed il mondo. Gli schemi e la scrittura sono più complessi dei concetti che corrispondono alle parole, tuttavia hanno delle similitudini nel fatto che consistono in una serie di caratteristiche che possono essere collegate ed applicate alle nuove situazioni. Lo schema esplicativo utilizzato nei sistemi basati sui concetti è il seguente:

Obiettivo della spiegazione:

— perché gli individui possiedono un particolare tipo di comportamento intelligente?

Concetto esplicativo:

— gli individui possiedono una serie di concetti, organizzati tramite delle aperture, che stabiliscono le tipologie gerarchiche ed altre associazioni;

— gli individui possiedono una serie di procedure finalizzate alle applicazioni concettuali, inclusi l'attivazione diffusa (*spreading activation*), il controllo e l'eredità;

— le procedure applicate ai concetti producono il comportamento;

— i concetti possono essere tradotti in regole, tuttavia possono differenziare le informazioni contrariamente ad una serie

di regole, dando luogo ad una serie di procedure computazionali differenti.

4.4 Analogie

Le analogie giocano un ruolo importante all'interno del pensiero umano, in aree distinte come il risolvere problemi, il prendere decisioni o la spiegazione e la comunicazione linguistica. I modelli computazionali simulano il modo in cui gli individui attuano i processi di recupero e mappano le fonti analogiche in modo da applicarle nelle situazioni tipo. Lo schema esplicativo per le analogie è:

Obiettivo della spiegazione:

— perché gli individui possiedono una particolare tipologia di comportamento intelligente?

Concetto esplicativo:

— gli individui possiedono delle rappresentazioni verbali e visive relative alle situazioni che possono essere utilizzate come casi o analogie;

— gli individui possiedono processi di recupero, progettazione o adattamento che operano in queste analogie;

— i processi analogici, applicati alle rappresentazioni delle analogie, producono il comportamento.

Le limitazioni della somiglianza, della struttura e della decisione superano il difficile problema di come le esperienze precedenti possono essere ricordate e utilizzate per risolvere il nuovo problema. Non tutto il pensiero è analogico, e usando analogie inappropriate il pensiero può essere ostacolato, tuttavia le analogie possono risultare efficaci in applicazioni come l'educazione ed il disegno.

Paul Thagard

4.5 Immagini

Qualunque tipo di immagine, incluse quelle visive, gioca un ruolo importante in relazione al pensiero umano. Le illustrazioni catturano le informazioni visive e spaziali in un modo decisamente più utile rispetto alla descrizione verbale. Le procedure computazionali si adattano bene alle rappresentazioni visive, incluse l'ispezione, la ricerca, la focalizzazione, la rotazione e la trasformazione. Queste operazioni possono risultare utili a generare piani e spiegazioni in campi nei quali vengono applicate le illustrazioni. Lo schema esplicativo delle rappresentazioni visive è:

Obiettivo della spiegazione:

— perché gli individui possiedono un particolare tipo di comportamento intelligente?

Concetto esplicativo:

— gli individui possiedono le immagini visive delle situazioni;
— gli individui possiedono processi mentali come la scansione (*scanning*) e la rotazione che operano in queste immagini;
— i processi relativi alla creazione e manipolazione delle immagini producono il comportamento intelligente.

Il linguaggio figurato comporta l'apprendimento e alcuni aspetti metaforici del linguaggio hanno origine nel linguaggio figurato. Gli esperimenti psicologici suggeriscono che le procedure visive come la scansione e la rotazione utilizzano il linguaggio figurato e risultati neurofisiologici recenti confermano l'esistenza di una correlazione fisica tra il ragionamento, le immagini mentali e la percezione.

4.6 Connessioni neurali

Le reti connessionistiche di semplici nodi o anelli sono molto utili per capire i processi psicologici che coinvolgono la

soddisfazione costrittiva parallela. Tali processi includono aspetti della visione, decisioni, selezione di spiegazione e significato nella comprensione del linguaggio. I modelli connessionistici possono simulare l'apprendimento con metodi che includono l'apprendimento di Hebbian e la retro-propagazione. Questo schema spiega come avviene questo approccio connessionistico:

Obiettivo della spiegazione:

— perché le persone hanno alcuni tipi di comportamento mentali?

Concetto esplicativo:

— le persone hanno rappresentazioni che coinvolgono semplici unità di processo legate tra loro da connessioni eccitatorie o inibitorie;

— le persone hanno processi che sprigionano attivazione tra le unità attraverso le loro connessioni, così come hanno processi per modificare le connessioni;

— il comportamento nasce quindi dall'applicazione alle unità di apprendimento e attivazione in crescita.

Simulazioni di vari esperimenti psicologici hanno dimostrato l'importanza dei modelli connessionistici, che costituiscono comunque solo un'approssimazione della vera rete neurale. Negli'ultimi anni, i modelli computazionali del cervello sono diventati biologicamente più ricchi, da una parte impiegando neuroni più realistici, tra cui quelli con potenziale d'azione (*spiking*) compreso, e dall'altra simulando l'interazione tra le varie zone del cervello come l'ippocampo e la corteccia. Questi modelli non sono esattamente un'alternativa alle descrizioni computazionali in termini di logica, concetti, regole, immagini e connessioni, ma dovrebbero essere loro compatibili e mostrare come avviene a livello neurale il funzionamento della mente.

5. Rilevanza filosofica

Una parte della filosofia, in particolare quella naturalistica della mente, fa parte della scienza cognitiva. Ma questo campo della scienza cognitiva ha importanza nella filosofia sotto vari aspetti. Prima di tutto l'aspetto psicologico, computazionale e altri risultati della ricerca della scienza cognitiva hanno importanti potenziali applicazioni ai problemi tradizionali filosofici in epistemologia, metafisica ed etica. In secondo luogo, la scienza cognitiva può essere oggetto di critica filosofica, specie riguardo al tema centrale in cui si afferma che il pensiero è rappresentativo e computazionale. In terzo luogo, la scienza cognitiva può essere oggetto di ricerca nella filosofia della scienza, producendo riflessioni sul metodo ed i presupposti del progetto.

5.1 Applicazioni filosofiche

Molte ricerche filosofiche sono oggi di tipo naturalistico, considerando continua la ricerca filosofica con lavori empirici in campi come quello psicologico. Dal punto di vista naturalistico, la filosofia della mente è strettamente alleata al lavoro teorico e sperimentale della scienza cognitiva. Bisogna giungere a conclusioni metafisiche sulla natura della mente non attraverso una speculazione a priori ma attraverso una riflessione sugli sviluppi scientifici in campi come l'informatica e le neuroscienze. In maniera simile, l'epistemologia non è l'unico esercizio concettuale ma dipende e trae benefici da scoperte scientifiche sulla struttura mentale e apprendimento delle procedure. Persino l'etica trae benefici dalle conoscenze del pensiero morale per raggiungere questioni morali come la natura della riflessione di cos'è giusto o sbagliato. Goldman (1993) ci fornisce una lista concisa delle applicazioni della scienza cognitiva all'epistemologia, filosofia della scienza, filosofia della mente, metafisica ed etica. Ecco alcuni dei problemi filosofici per cui gli sviluppi in corso della scienza cognitiva sono d'importanza vitale.

— Istinto. Fino a che punto la conoscenza è innata o acquisita con l'esperienza? Il comportamento umano è conseguenza della natura o dell'educazione?

— Linguaggio del pensiero. Il cervello umano opera tramite un codice come il linguaggio, o tramite una struttura connessionistica più generale? Che rapporto esiste tra i modelli cognitivi simbolici che utilizzano regole e concetti, e i modelli simbolici che utilizzano le reti neurali?

— Immagini mentali. La mente umana pensa attraverso le immagini visive o di altro genere, o solamente tramite le rappresentazioni come il linguaggio?

— Psicologia popolare. Per un individuo comprendere quotidianamente altri individui consiste nel possedere una teoria mentale, o semplicemente nell'avere la capacità di assecondarli?

— Significato. In che modo le rappresentazioni mentali acquisiscono i significati o i contenuti mentali? Fino a che punto il significato di una rappresentazione dipende dalla sua relazione con altre rappresentazioni, la sua relazione con il mondo, e la sua relazione con gli altri esseri pensanti?

— Identità mente-cervello.

— Volontà. L'azione umana è libera o semplicemente causata dagli eventi cerebrali?

Sorgono ulteriori problemi filosofici se esaminiamo i presupposti degli approcci alla scienza cognitiva.

5.2 Critiche alle scienza cognitiva

La pretesa che la mente umana lavori attraverso rappresentazioni e computazioni è una congettura empirica, e potrebbe essere sbagliata. Nonostante l'approccio computazionale-rappresentativo alla scienza cognitiva sia servito a spiegare vari aspetti della risoluzione dei problemi umani – apprendimento, uso del linguaggio – alcuni critici filosofici come Hubert Dreyfus (1992) e John Searle (1992) ritengono

che questo tipo di approccio sia fundamentalmente sbagliato. I critici della scienza cognitiva riassumono queste sfide in:

- sfida emotiva: la scienza cognitiva trascura il ruolo importante che svolgono le emozioni nel pensiero umano;
- sfida della coscienza: la scienza cognitiva ignora l'importanza della coscienza nel pensiero umano;
- sfida del mondo: la scienza cognitiva trascura il significativo ruolo dell'ambiente fisico circostante nel pensiero umano;
- sfida del corpo: la scienza cognitiva trascura il contributo del corpo nel pensiero e nelle azioni umane;
- sfida sociale: la scienza cognitiva ignora quanto il pensiero umano sia inerente al sociale;
- sfida dei sistemi dinamici: la mente è un sistema dinamico e non un sistema computazionale;
- sfida matematica: risultati matematici dimostrano come il pensiero umano non può essere computazionale nel senso standard: il cervello opera diversamente, forse come un computer quantistico.

Thagard (1996) sostiene che si può andare incontro a tutte queste sfide ampliando ed integrando gli approcci computazionali-rappresentativi, ma non abbandonandoli.

5.3 Filosofia della scienza cognitiva

La scienza cognitiva suscita varie questioni metodologiche degne di ricerca da parte dei filosofi della scienza. Qual è la natura della rappresentazione? Che ruolo giocano i modelli computazionali nello sviluppo delle teorie cognitive? Che relazione esiste tra i valori della mente che coinvolgono l'elaborazione simbolica, le reti neurali ed i sistemi dinamici? Qual è il rapporto tra i campi della scienza cognitiva come la psicologia, la linguistica e la neuroscienza? I fenomeni psicologici sono soggetti a spiegazioni riduzionistiche attraverso la neuroscienza? Von Eckardt (1993) e Clark (2001) danno vita a discussioni su alcuni argomenti filosofici che sfociano nella

scienza cognitiva. Bechtel et al. (2001) raccolgono articoli sulla filosofia della neuroscienza.

Bibliografia

- BECHTEL, W., GRAHAM, G. (a cura di), *A companion to Cognitive Science*, Malden, Blackwell, 1998.
- BECHTEL, W., MANDIK, P., MUNDALE, J., STUFFLEBEAM, R. S. (a cura di), *Philosophy and Neurosciences, A Reader*, Malden, Blackwell, 2001.
- CLARK, A., *Mindware: An Introduction to the Philosophy of Cognitive science*, New York, Oxford University Press, 2001.
- DAWSON, M. R. W., *Understanding Cognitive Science*. Oxford, Blackwell, 1998.
- DREYFUS, H. L., *What Computers Still Can't Do*. (3rd ed.). Cambridge, MIT Press, 1992.
- ELIASMITH, C., Anderson, C.H., *Neural Engineering: Computation, Representation and Dynamics in Neurobiological Systems*. Cambridge, MIT Press, 2003.
- GOLDMAN, A., *Philosophical Applications of Cognitive Science*. Boulder, Westview Press, 1993.
- JOHNSON-LAIRD, P., *The Computer and the Mind: An Introduction to Cognitive Science*, Cambridge, Harvard University Press, 1998.
- NADEL, L. (a cura di), *Encyclopedia of Cognitive Science*. London, Nature Publishing Group, 2003.
- POLK, T. A., SEIFERT, C.M. (a cura di), *Cognitive Modeling*. Cambridge, MIT Press, 2002.
- SEARLE, J., *The Rediscovery of the Mind*. Cambridge, MIT Press, 1992.
- SOBEL, C. P., *The Cognitive Sciences: An Interdisciplinary Approach* Mountain View, Mayfield, 2001.
- STILLINGS, N., et al., *Cognitive Science*, Cambridge, MIT Press, 1995.
- THAGARD, P., *Mind: Introduction to Cognitive Science*, Cambridge, MIT Press (seconda edizione), 1996.

Paul Thagard

VON ECKARDT, B., *What is Cognitive Science?* Cambridge, MIT Press, 1993.

WILSON, R.A., Keil, F.C. (a cura di), *The MIT Encyclopedia of the Cognitive Sciences*. Cambridge, MIT Press, 1999.

Ringraziamenti

Con il gentile permesso del MIT Press, questo articolo accorpa materiali tratti dalla prima e dalla seconda edizione di P. Thagard, *Mind: Introduction to Cognitive Science*.

LA TEORIA DELL'IDENTITÀ DELLA MENTE⁴⁵

J. J.C. Smart

La teoria dell'identità della mente afferma che stati e processi mentali sono identici a stati e processi cerebrali. Questo non implica che mente e cervello debbano essere considerati identici. Le espressioni «che mente che ha!», e «che cervello che ha!», vengono usate comunemente in modo intercambiabile, ma non si direbbe mai «la sua mente pesa un chilo e mezzo». Un'analisi della mente e del cervello diviene quindi un'analisi dei processi e, forse, degli stati mentali e cerebrali. Consideriamo per esempio l'esperienza del dolore, del vedere qualcosa, oppure dell'avere un'immagine mentale. La teoria dell'identità sostiene che tali esperienze non sono solo semplicemente correlate a processi mentali, bensì costituiscono veri e propri processi mentali.

Alcuni filosofi, pur riconoscendo che le esperienze sono processi cerebrali, ritengono che esse posseggano proprietà psichiche e non fisiche, denominate «*qualia*». La teoria dell'identità nega l'esistenza di tali irriducibili caratteristiche non fisiche. Alcuni teorici dell'identità compiono un'analisi comportamentista degli stati mentali, come le credenze ed i desideri, mentre altri, chiamati a volte materialisti dello stato centrale (*central state materialists*), ritengono che gli stati mentali siano dei veri e propri stati cerebrali. I teorici

⁴⁵ J.C.C. Smart, *The Identity Theory of Mind*, in «Stanford Encyclopedia of Philosophy», <http://plato.stanford.edu/entries/mind-identity/>

dell'identità spesso si autodefiniscono «materialisti», ma il termine «fisicalisti» sarebbe forse più appropriato. In effetti si potrebbe avere un approccio materialista rispetto alla mente, riconoscendo nondimeno che vi sono degli elementi della fisica che non sarebbero definiti adeguatamente col termine «materiali».

Se si considera la teoria dell'identità (nelle sue svariate forme) come una sorta di fisicalismo, la si dovrà intendere come un fisicalismo ontologico piuttosto che come un fisicalismo «di traduzione». Sarebbe del tutto assurdo tentare di tradurre frasi contenenti la parola «cervello» o «sensazione» in frasi riguardanti elettroni, protoni, e così via. Ancor più irragionevole, del resto, apparirebbe il tradurre ad esempio frasi contenenti la parola «albero». Dopo tutto il termine «albero» è appreso in modo ampiamente ostensivo, e non rientra nemmeno nella classificazione botanica. Se fossimo sufficientemente piccoli, un dente di leone potrebbe apparirci come un albero. Tuttavia un fisicalista potrebbe dire che gli alberi sono complicati meccanismi fisici. Il fisicalista rifiuterà il concetto forte di emergenza, nel senso di filosofi come Samuel Alexander e C.D. Broad. Quest'ultimo osservò (Broad, 1937) che in base alle conoscenze dell'epoca era impossibile desumere le proprietà del sale da quelle del sodio e del cloro presi singolarmente (sviluppiò lo stesso discorso anche in termini epistemologici: la teoria del caos è un esempio di come, anche in una teoria deterministica, le conseguenze fisiche possono superare le previsioni). Naturalmente il fisicalista non rifiuterà il senso innocuo del termine «emergenza», per il quale un apparato non è solamente un insieme confuso delle sue parti (Smart, 1981).

1. *Antecedenti storici*

La teoria dell'identità, come è qui intesa, risale agli anni '50, al lavoro degli studiosi U. T. Place e Herbert Feigl. Nel corso della storia, diversi filosofi e scienziati hanno abbracciato il materialismo; tra di essi Leucippo, Hobbes, La Mettrie, d'Holbach e Karl Vogt, il quale fece la paradossale osserva-

zione (forse da non prendersi troppo alla lettera) secondo cui il cervello secernerebbe i pensieri nello stesso modo in cui il fegato secerne la bile. Sarà tuttavia grazie al saggio di U.T. Place *Is Consciousness a Brain Process?* (Place, 1956) e a quello di H. Feigl intitolato *The 'Mental' and the 'Physical'* (Feigl, 1958), che la teoria dell'identità comincerà a suscitare l'interesse degli studiosi. A tal proposito vanno menzionati i contributi di Rudolf Carnap (1932, pag. 127), H. Reichenbach (1938) e M. Schlick (1935). Secondo Reichenbach gli eventi mentali possono essere identificati attraverso i corrispettivi stimoli e reazioni, così come lo stato interno di una cellula fotoelettrica (probabilmente sconosciuto) può essere identificato dallo stimolo (la luce) e dalla relativa reazione (il flusso di corrente elettrica). In entrambi i casi, gli stati interni possono essere stati fisici. Carnap considerava l'identità nel senso di un consiglio linguistico, piuttosto che come qualcosa di positivamente connesso a una questione di fatto. In merito a questo concetto si veda il suo saggio *Herbert Feigl on Physicalism* (in Schilpp, 1963, cfr. in partic. pag. 886). Fu probabilmente lo psicologo E. G. Boring (1933) il primo ad utilizzare il termine «teoria dell'identità» (v. Place, 1990).

Il saggio di Place, originale e pionieristico, fu scritto in seguito a numerose discussioni avvenute con J. J. C. Smart e C. B. Martin all'Università di Adelaide. Per una raccolta dei contributi di Martin vedi Place (1989) *Low Claim Assertions in Heil* (1989). A quell'epoca Smart appoggiava una posizione comportamentista, per la quale gli eventi mentali venivano intesi in termini di pure asserzioni ipotetiche sul comportamento, al pari di resoconti di esperienze in prima persona, quelle che Gilbert Ryle definiva «dichiarazioni»⁴⁶. Le dichiarazioni venivano intese come semplici parti del comportamento, come se il dire di sentire un dolore costituisse solamente una forma sofisticata di «sussulto» (*wince*). Smart riteneva che la teoria di Ryle fosse compatibile con il fisicalismo, malgrado questo non rientrasse nelle intenzioni di Ryle.

⁴⁶ Il termine è *Avowals*, che letteralmente significa «dichiarazioni», «ammissioni», «confessioni pubbliche» (Ndt).

Smart si augurava che gli ipotetici potessero essere spiegati dalla neuroscienza e dalla cibernetica. Non potendo rifiutare la posizione di Place, e ritenendo allo stesso tempo insoddisfacente il modo in cui Ryle trattava le esperienze interne, cosa che lo stesso Ryle riconobbe in parte (Ryle 1949, pag. 240), Smart si convertì presto alla posizione presa da Place (Smart 1959). In questo lo scritto di Feigl *The 'Mental' and the 'Physical'* (Feigl 1958, 1967) giocò un ruolo importante. Il vasto contributo di Feigl trattava molte questioni, incluse quelle relative all'intenzionalità, introducendo peraltro l'utile termine «*nomological danglers*»⁴⁷ per le correlazioni tra fisico e mentale, così come erano intese dai dualisti. Esse «penzolarebbero» (*dangle*), appunto, dalla rete nomologica della fisica e colpirebbero l'attenzione come implausibili escrescenze sul volto immacolato della scienza. Il saggio di Feigl (1967) contiene peraltro anche un rimarchevole «poscritto».

2. La natura della teoria dell'identità

Place parlava di costituzione, piuttosto che di identità. Tra i suoi esempi gli enunciati «questo tavolo è una vecchia cassa da imballaggio», e «il lampo è una scarica elettrica». A quest'ultimo enunciato, in particolare, Place aveva già fatto cenno in un suo precedente saggio *The Concept of Heed* (Place, 1954), in cui aveva ripreso il comportamentismo di Ryle così come si applicava ai concetti di coscienza, sensazione e immaginazione mentale. Place osservò:

Le logiche obiezioni che potrebbero essere mosse rispetto ad un'affermazione come «la coscienza è un processo del cervello» non sono maggiori di quelle che potrebbe suscitare l'affermazione «il lampo è una scarica elettrica».

⁴⁷ Espressione intraducibile in italiano: letteralmente «pendenti nomologici», «appendici nomologiche pendenti», «escrescenze nomologiche»: *dangle* vuol dire infatti oscillare, penzolare, dondolare e far balenare (*ndt*).

Va notato che il significato attribuito da Place alla parola «logico» era quello in voga ad Oxford a quell'epoca e non quello attribuitogli comunemente oggi. Un'obiezione riguardava il fatto che «sensazione» non è la stessa cosa che «processo cerebrale». In risposta a tale obiezione Place fece presente che «questo tavolo» non ha lo stesso significato di «questa vecchia cassa da imballaggio», e che «lampo» non ha lo stesso significato di «scarica elettrica». Le operazioni implicate nel determinare la natura di un tavolo sono diverse da quelle impiegate per determinare la natura di una vecchia cassa da imballaggio. Ci accorgiamo di un lampo attraverso lo sguardo, ma che si tratti del movimento di cariche elettriche lo comprendiamo attraverso la teoria e lo studio. Questo, d'altronde, non impedisce al tavolo di essere identico alla vecchia cassa da imballaggio, e al lampo osservato di non essere altro che una scarica elettrica. Feigl e Smart, invece, posero il problema in termini di distinzione tra significato e referente. «Sensazione» e «processo cerebrale» possono avere significati diversi ma si riferiscono alla stessa cosa. Le espressioni: «un pianeta molto brillante visto di giorno», e: «un pianeta molto brillante visto di sera», fanno ambedue riferimento alla stesso corpo celeste, Venere. (Ovviamente queste espressioni potrebbero essere costruite per riferirsi a cose differenti, a differenti sequenze delle fasi temporali di Venere, ma non necessariamente o automaticamente si deve intendere così).

Vi era stata fino ad allora tra i filosofi la tendenza a ritenere che le affermazioni sull'identità fossero necessarie e costituissero della verità a priori. Tuttavia i teorici dell'identità consideravano l'enunciato 'le sensazioni sono processi cerebrali' come contingente. Bisognava ancora scoprire che il concetto di identità avrebbe resistito. Dopo tutto, Aristotele pensava che il cervello servisse a raffreddare il sangue. Cartesio, che la coscienza fosse immateriale.

Ci si è opposti talvolta all'idea che gli enunciati sulle sensazioni siano non correggibili (*incorrigible*), mentre quelli sul cervello siano correggibili (*corrigible*), avanzando l'ipotesi che le sensazioni debbano essere in qualche modo differenti. Sia

Ryle che Wittgenstein sostennero l'affascinante ma poco plausibile teoria per cui gli apparenti resoconti delle esperienze immediate non sarebbero in realtà resoconti, bensì dichiarazioni (*avowals*), come se il mio resoconto di un mal di denti fosse una sofisticata forma di sussulto (*wince*). Influenzato da Martin, Place poté spiegare la relativa incorreggibilità degli enunciati sulle sensazioni attraverso le loro asserzioni deboli: «vedo un remo piegato» è una asserzione più forte di «mi sembra che ci sia un remo piegato». Ad ogni modo, la mia sensazione e la mia apparente consapevolezza della sensazione costituiscono due diverse entità, e quindi, secondo il principio di Hume, deve essere possibile che ce ne possa essere una senza l'altra. Bisognerebbe rifiutare tutto fuorché una incorreggibilità relativa (Place, 1989). Come già ribadito, Place preferiva esprimere la teoria attraverso la nozione di costituzione, mentre Smart era più incline a sottolineare il concetto d'identità così come viene espresso, in logica, dagli assiomi dell'identità. Per Smart, se la sensazione X è identica al processo cerebrale Y, e se Y si trova tra le mie orecchie ed è retto o circolare, allora la sensazione X si trova tra le mie orecchie, ed è di forma retta o circolare. Naturalmente non ci apparirebbe così nell'esperienza immediata. Solo il neuroscienziato, forse, potrebbe conoscere la sua forma. Per esemplificare: un professore di anatomia potrebbe essere anche il rettore della stessa facoltà di medicina; uno studente potrebbe sapere che il professore (di anatomia) ha il singhiozzo durante le lezioni senza necessariamente sapere che anche al rettore succede la stessa cosa.

3. *Proprietà fenomenali e analisi topico-neutrali*

Qualcuno potrebbe pensare che il rettore di una scuola di medicina, in quanto tale, non insegni. Egli, in quanto decano, si riunisce con il vice-cancelliere. Non è questa la cosa fondamentale, ma c'è una cosa fondamentale dietro di essa. La nostra analisi si sofferma sul fatto che le proprietà dell'essere professore di anatomia non sono identiche alle proprietà dell'essere il rettore di una scuola medica. A questo punto

una domanda sorge spontanea: anche se le sensazioni sono identiche ai processi cerebrali, non esistono forse delle proprietà non-fisiche introspettive di sensazioni che differiscono dalle proprietà dei processi cerebrali? La risposta (Smart, 1959) è che le proprietà delle esperienze sono «topico-neutrali». Smart ha tratto l'espressione «topico-neutrale» da Ryle, che a sua volta la ha utilizzata per designare termini come «se», «o», «e», «no», «perché». Se riconosciamo tali elementi isolatamente all'interno di una conversazione, non siamo in grado di capire se la conversazione verta sulla matematica, sulla fisica, sulla geologia, sulla storia, sulla teologia, o su qualsiasi altra materia. Smart ha utilizzato l'espressione «topico-neutrale», nel significato più prossimo all'essere neutrale tra fisicalismo e dualismo. Ad esempio «continuare», «accadere», «intermittente», «crescere», «decrescere», sono topico-neutrali. Lo è anche il termine «me», in quanto riferito a chi formula la frase in questione. Pertanto affermare che una sensazione è causata da un fulmine o dalla vista di una rapa non assicura che la sensazione sia non-fisica, a parere dei dualisti, o sia fisica, secondo i materialisti. Questa frase è inoltre neutrale, sia che le proprietà siano fisiche, sia che alcune di esse siano irriducibilmente psichiche. Per capire come questa idea possa essere applicata al presente scopo, facciamo un esempio. Supponiamo di avere un'immagine mentale a strisce gialle verdi e viola. A questo punto potremmo introdurre il termine filosofico «*datum* sensibile» (*sense datum*) per chiarire il concetto di visione o di presunta visione di qualcosa di verde giallo e viola: noi diciamo di avere un *datum* sensibile del verde, del giallo e del viola. Cioè, noi vedremmo o ci sembrerebbe di vedere, per esempio, una bandiera o dei lampi a strisce verdi gialle e viola. Supponiamo inoltre che, come sembra plausibile, non ci sia niente a strisce gialle verdi e viola nella nostra mente. Si deduce quanto sia importante per i teorici dell'identità affermare (come in effetti è stato fatto) che i *data* sensibili e le immagini non fanno parte del contenuto del mondo. L'affermazione «io ho un *datum* sensibile del verde» significa in realtà che io vedo, o mi sembra di vedere, qualcosa di verde. Questa mossa non dovrebbe esser vi-

sta come un mero stratagemma *ad hoc*, dato che Ryle e J.L. Austin, nonché Wittgenstein ed altri, ne hanno fornito le argomentazioni. Si pensi all'affermazione di Ryle, secondo cui le immagini mentali non sono una sorta di cartolina illustrata con immagini spettrali. Place ha caratterizzato la «fallacia fenomenologica» come l'errore nel pensare che quando noi percepiamo qualcosa di verde stiamo percependo qualcosa di verde nella mente. Egli caratterizza tale fallacia come:

l'errore di supporre che quando il soggetto descrive la sua esperienza, quando egli descrive le cose in base all'apparenza, al loro suono, odore, sapore, o alle sensazioni che tali cose gli suscitano, egli non fa altro che descrivere le proprietà letterali degli oggetti e degli eventi su una sorta di peculiare schermo cinematografico o televisivo interno, generalmente definito nella letteratura psicologica moderna «campo fenomenale».

Ovviamente, come lo stesso Smart riconobbe, ciò fa dipendere la teoria dell'identità da una considerazione physicalista del colore. La nozione di colore proposta inizialmente da Smart risultò essere troppo comportamentista, e incapace di affrontare, ad esempio, il problema dello spettro inverso; in seguito egli sviluppò una concezione realista e oggettivista. Armstrong è stato realista circa il colore, ma Smart temeva che in tale prospettiva quello di colore sarebbe stato un concetto molto disgiuntivo e idiosincratico, di nessuna importanza «universale», di nessun interesse, ad esempio, per extraterrestri dotati di differenti sistemi visivi. Traendo ispirazione da una conversazione con Lewis, Smart concluse che questa non era un'obiezione valida all'idea che i colori fossero delle proprietà oggettive.

Per prima cosa va spiegato come un normale individuo percepisce i colori. A tal fine esistono test obiettivi, in termini di abilità, basati sulle discriminazioni dei colori. Ciò può essere fatto senza circolarità. Quindi «discriminare i colori» è una nozione più primitiva di quella del colore stesso. (Confronta il modo in cui nella teoria degli insiemi «equinumeroso» è antecedente a «numero»). In seguito Smart ha chiarito la no-

zione del colore in termini di discriminazione di esso da parte di soggetti umani medi, che percepiscono in condizioni normali (come dire un'alba scozzese nuvolosa). Tale nozione di colore potrebbe essere disgiuntiva e idiosincratia (le equazioni di Maxwell potrebbero servire per studiare gli Alfa Centauri, ma difficilmente lo sono le nostre nozioni di colore). Per quanto tali criteri siano antropocentrici e disgiuntivi, sono tuttavia anche oggettivi. David R. Hilbert (1987) paragona i colori alle riflettanze, riducendo così la disgiuntività e l'idiosincrasia. Pochi epicicli sono facilmente aggiunti per affrontare la luce radiata, i colori dell'arcobaleno, o il sole al tramonto, o i colori prodotti dalle diffrazioni provenienti dalla piume.

John Locke imboccò la strada giusta considerando le qualità secondarie oggettive come poteri dell'oggetto, ma sbagliò ritenendo tali poteri capaci di produrre idee nella mente piuttosto che di generare discriminazioni comportamentali. Anche Smart direbbe che se tali poteri sono delle disposizioni, allora dovremmo considerare le qualità secondarie come basi categoriche di tali poteri (ad esempio, nel caso delle proprietà dei colori delle superfici degli oggetti). La teoria di Locke si basa sul fatto che le idee possiedano *qualia* misteriosi, osservati sulla scena di un interno teatro mentale. In ogni caso, per rendere giustizia a Locke, va detto che egli non parla effettivamente di «idee rosse» ma di «idee del rosso». I filosofi che intendono «è rosso» nel senso di «appare rosso» affrontano il problema nel modo sbagliato.

Torniamo ora alla questione di partenza. Cosa significa avere un *datum* sensibile o un'immagine mentale a strisce gialle, verdi e viola, posto che non c'è nessun oggetto giallo verde e viola nella nostra mente? I teorici dell'identità possono affermare che i *data* sensibili e le immagini non esistono nella realtà: più o meno come un idraulico medio. Frasi relative ad un idraulico generico possono essere estese all'intera categoria. Allo stesso modo, avere un *datum* sensibile o un'immagine mentale del verde non è avere il *datum* sensibile; così come avere un *datum* sensibile o un'immagine menta-

le del verde non è l'aver il verde stesso. Il processo cerebrale non è (il) verde. Scrive Place (1956, pag. 49):

Quando noi descriviamo l'immagine mentale residuale (*after-image*) come verde [...], diciamo che stiamo facendo il tipo di esperienza che normalmente facciamo, e che abbiamo imparato a descrivere guardando una macchia verde di luce.

E Smart (1959) afferma:

Quando una persona dice «io vedo un'immagine residuale (*after-image*) arancio-giallastra», sto dicendo più o meno che «succede qualcosa che somiglia a ciò che succede quando ho gli occhi aperti, sono cosciente e c'è dell'arancio ben illuminato di fronte a me».

Riprendendo tali citazioni, David Chalmers (1996, pag. 360) afferma che se la frase «succede qualcosa» è elaborata in senso sufficientemente ampio, allora essa non è adeguata, se invece è elaborata in modo specifico e si riferisce solo agli stati (o processi) dell'esperienza non è ugualmente sufficiente per una conclusione. Smart lo contraddirebbe, utilizzando con particolare evidenza il termine «proprio». Più esplicitamente: una serie di cose succedono in me quando ho un'immagine residuale gialla (ad esempio il mio cuore sta pompando sangue che giunge al cervello); tuttavia, tali cose avvengono a prescindere, e non sono *proprie* del momento. Contro Place, Chalmers sostiene che la parola «esperienza» non è analizzabile, e così l'analisi di Place non è sufficiente per stabilire un'identità tra sensazioni e processi cerebrali. Sempre contro Smart, egli sostiene che lasciare la parola «esperienza» fuori dall'analisi rende la stessa inadeguata. Egli in altre parole non accetta l'analisi «topico-neutrale». In difesa di Place, si potrebbe affermare, forse riprendendo quanto sostiene Farrel (1950), che non è chiaro che non si possa fornire un'analisi della parola «esperienza». Se non abbiamo bisogno della parola «esperienza», allora non abbiamo bisogno neanche della parola «mentale». Rosenthal sostiene (contro

i teorici dell'identità) che le esperienze hanno alcune proprietà tipicamente mentali, e che «inevitabilmente perdiamo ciò che contraddistingue il mentale se analizziamo queste proprietà come né fisiche né mentali». In effetti, essere topico-neutrale significa poter essere tanto fisico quanto mentale, così come lo è l'aritmetico. Non è necessario che la parola «mentale» compaia nella formula topico-neutrale. «Mentale», a parere di Ryle (1949), nel suo uso ordinario è piuttosto un termine polivalente, «aritmetico mentale», «malattia mentale», ecc... con cui i teorici dell'identità eliminano ogni dubbio.

4. *Le teorie del ruolo causale*

Nei loro studi sulla mente, David Lewis e D. M. Armstrong evidenziano il concetto di causalità. In una sua chiara esposizione della teoria dell'identità, Lewis (1966) scrive (cito il passo da Lewis, 1983, pag. 100):

La caratteristica principale di ogni (tipo di) esperienza come tale è il suo ruolo causale, il suo insieme di cause ed effetti più tipici. Noi materialisti, invece, crediamo che tali ruoli causali che appartengono, per necessità analitica, alle esperienze, appartengano, in realtà, a certi stati fisici. Dal momento, poi, che questi stati fisici possiedono le caratteristiche delle esperienze, essi debbono essere, necessariamente, esperienze.

Sulla stessa scia, Robert Kirk (1999) propone una tesi sull'impossibilità dell'esistenza degli zombie. Se il supposto zombie possedesse tutte le caratteristiche comportamentali e mentali ascrittegli da coloro che propugnano l'esistenza degli zombie contro il materialismo, lo zombie diventerebbe cosciente e, quindi, cesserebbe di essere uno zombie.

Non vi è, quindi, la necessità di un uso esplicito del Rasoio di Ockham, come visto in Smart (1959) – ma non in Place (1956; v. anche Place 1960). L'importante saggio di Lewis proponeva già un'unione tra la teoria dell'identità della mente e le così dette idee «funzionaliste», rese esplicite da Lewis

stesso (1972 e 1994). Nel testo del 1972, intitolato *Psychophysical and Theoretical Identifications*, Lewis utilizza alcuni dei concetti del suo saggio più formale, *How to Define Theoretical Terms* (1970). La psicologia popolare *folk psychology* contiene vocaboli che definiamo «psicologici», come «sensazione», «percepire», «credenza», «desiderio», «emozione», ecc., così come sono ricorrenti i termini associati a colori, odori, suoni e sapori. Si potrebbe considerare una serie di banalità del senso comune contenenti entrambe queste categorie di parole come una teoria, e potremmo poi raccogliere termini teorici di una psicologia del senso comune, evidenziando ogni entità o ogni tipo di entità che realizza perfettamente la teoria. Se certi stati neurali agiscono nello stesso modo (come noi crediamo), allora gli stati mentali devono essere questi stessi stati neurali. Nel suo saggio del 1994, Lewis si basa sul tatto per ricavare una teoria direttamente dal senso comune. Non si possono affastellare acriticamente le banalità implicite nei nostri modelli linguistici, come si fa nella creazione di una grammatica, bisogna prendere le mosse da ciò che, nella nostra migliore teoria, costituirebbe la grammaticalità.

Un grande vantaggio di questo approccio, rispetto alla prima teoria dell'identità, è l'olismo. In particolare, sono due gli aspetti che vanno presi in considerazione. Anzitutto tale approccio è in grado di spiegare le interazioni causali tra gli stati cerebrali ed i processi stessi, così come nel caso degli stimoli esterni e delle risposte. In secondo luogo, tale orientamento fa riferimento alla nozione di «ramseificazione» di una teoria. F. P. Ramsey ha mostrato come sostituire la terminologia di una teoria, come per esempio, «la proprietà di un elettrone...» con «la proprietà X tale che», cosicché, quando questa sostituzione sarà avvenuta per tutti i termini teorici, non ci rimarrà altro che «proprietà X tale che», «proprietà Y tale che» ecc. Consideriamo, per esempio, i termini che descrivono il comportamento come termini di osservazione, ed i termini psicologici come termini della psicologia popolare. La «ramseificazione» mostra come la psicologia popolare sia compatibile con il materialismo. Tutto ciò appare corretto, anche se le prime forme della teoria dell'identità af-

frontano in modo più diretto i resoconti delle esperienze immediate.

L'approccio causale fu anche una caratteristica dell'attenta analisi concettuale degli stati e processi mentali, come la percezione, e delle qualità secondarie, sensazione, coscienza, credenza, desiderio, emozione e azione volontaria che D. M. Armstrong rese nota nel suo *A Materialist Theory of the Mind* (1968a), e poi in una seconda edizione (1993), arricchita da una nuova prefazione. La prima e la seconda parte di questo libro sviluppano un'analisi concettuale che spiana la strada all'identificazione contingente degli stati e dei processi mentali con quelli materiali. Come fece Brian Medlin, in una notevole critica a Ryle in difesa del materialismo (Medlin 1967), Armstrong preferì classificare la teoria dell'identità in termini di «materialismo dello stato centrale» (Central State Materialism). Indipendentemente da Armstrong e Lewis, ma in modo analogo, il materialismo dello stato centrale di Medlin si incentrava su un'analisi causale dei concetti riguardanti gli stati ed i processi mentali. Vedi Medlin 1967 e 1969 (inclusa nota 1).

Sarebbe opportuno menzionare, in questa sede, altri due libri di Armstrong, il primo sulla percezione (1961), e l'altro sulle sensazioni corporee (1962). Per Armstrong percepire consisteva nel giungere a credere qualcosa attraverso i sensi (v. anche Pitcher, 1971). Questa prospettiva associa i vantaggi del realismo diretto (*Direct Realism*) alla possibilità di accogliere la tendenza scientifica causale, che si riteneva essere d'appoggio alla precedente teoria della percezione. Armstrong trattava le sensazioni corporee come percezioni degli stati del nostro corpo. Naturalmente questi ultimi si sarebbero potuti confondere con gli stati emozionali, in quanto un prurito potrebbe includere una propensione a grattarsi oppure, come in rare circostanze, il dolore potrebbe essere sentito senza sofferenza. In ogni caso Armstrong riteneva che fosse la percezione la nozione centrale – la qual cosa presenta un problema terminologico. Smart aveva parlato di sensazioni visive, che non erano da considerarsi come percezioni, bensì come qualcosa che accade nella percezione. Intendendo in

tal modo le «sensazioni», dovrebbero esistere sensazioni di sensazione corporea. L'ambiguità potrebbe essere risolta utilizzando la parola «sentire» (*sensing*) nel contesto del «visivo», «uditivo», «tattile» e «corporeo», così che le sensazioni corporee sarebbero percezioni che riguardano «forme di sentire» introspettive. Dal momento che queste sensazioni corporee sono delle percezioni, si potrebbero verificare anche alcuni errori, come quando una persona con un piede amputato crede di sentire dolore proprio a quel piede. Egli ha il senso di «sentire dolore al piede», ma la parola non contiene il dolore al piede, così come non contiene i dati sensibili o immagini, ma contiene l'aver dati sensibili e rappresentazioni come immagini.

Il materialismo dello stato centrale di Armstrong riguardava l'identificazione di credenze e desideri con gli stati cerebrali (1968a). Mentre Smart finì per accettare questa idea, Place si oppose alla possibilità di estendere la teoria dell'identità anche agli stati delle disposizioni comportamentali come le credenze ed i desideri, sottolineando il fatto che noi non abbiamo accesso diretto alle nostre credenze e desideri. Come Ryle, Place riteneva che questi potessero essere spiegati in termini di enunciati ipotetici sul comportamento, usando l'analogia dei cavalli vapore di un'automobile (Place 1967). Il problema, qui, non risiede tanto nella base neurofisica degli stati mentali, quanto invece nella natura delle disposizioni comportamentali. La sua opinione rispetto a queste ultime è largamente discussa nel dibattito con Armstrong e Martin (cfr. il testo a cura di Armstrong, Martin e Place, T. Crane, 1996). Che gli stati mentali, come le credenze ed i desideri, siano disposizioni comportamentali oppure stati neurofisiologici descritti come topico-neutrali non deve preoccuparci al momento, mentre è importante tornare su quello che sembra l'elemento più complesso della coscienza. Le teorie dell'identità causale (*causal identity theories*) sono strettamente correlate al funzionalismo – aspetto che verrà trattato nel prossimo paragrafo. Smart era stato molto cauto nell'uso della nozione di causalità in metafisica, ritenendo che tale nozione non trovasse spazio nella fisica teorica. Smart avrebbe co-

munque dovuto riconoscere questo concetto nella psicologia popolare, come anche nella psicologia scientifica e nella biologia in generale, in cui la fisica e la chimica vengono applicate per spiegare generalizzazioni, piuttosto che leggi precise. Se la psicologia popolare fa uso della nozione di causalità, non ha importanza se si riferisca a quello che Quine chiama 'discorso di secondo grado', coinvolgendo le nozioni contestuali di modalità.

5. *Funzionalismo e teoria dell'identità*

Si è comunemente pensato che la teoria dell'identità sia stata rimpiazzata da una teoria denominata «funzionalismo». Si potrebbe dire che i funzionalisti esagerino molto la distanza tra il loro punto di vista e quello dei teorici dell'identità. In effetti alcuni filosofi come Lewis (1972 e 1994), Jackson, Parigger e Prior (1982), hanno ritenuto che il funzionalismo possa costituire una via in direzione della teoria dell'identità.

Allo stesso modo di Lewis e Armstrong, i funzionalisti definiscono gli stati ed i processi mentali sul piano delle loro relazioni causali con il comportamento, ma falliscono nell'identificarli con le loro realizzazioni a livello neurale. Naturalmente il termine «funzionalismo» è stato impiegato in maniere a volte vaghe e diverse, al punto che si potrebbe sostenere che persino le teorie di Place, Smart e Armstrong siano, in fondo, teorie funzionaliste. Nella parola «funzionalista» si riscontrano alcune affinità con la parola «funzione», così come viene intesa sia in biologia sia in matematica. In matematica una funzione è una serie ordinata di n -dati. In modo analogo, se i processi mentali sono definiti direttamente o indirettamente da una serie di coppie stimolo-risposta, le definizioni potrebbero essere viste come «funzionali» nel senso matematico. Pare probabile, comunque, che vi sia una relazione più stretta con l'accezione del termine in uso nelle scienze biologiche. In effetti si potrebbe definire la parola «occhio» in base alle sue funzioni anche se l'occhio di una mosca e quello di un cane sono anatomicamente e fisiologicamente assai differenti. Il funzionalismo identifica i processi

e gli stati mentali attraverso i loro ruoli di causalità e, come è stato notato in precedenza riguardo a Lewis, sappiamo che i ruoli funzionali sono posseduti dagli stati e dai processi neurali (ci sono peraltro anche delle forme teologiche e «homunculari» di funzionalismo; qui non ce ne occuperemo). In ogni modo un dualista interazionista come il famoso neurofisiologo Sir John Eccles negherebbe (non plausibilmente, per la maggior parte di noi) che tutti questi ruoli funzionali siano in realtà posseduti. Si potrebbe pensare alla psicologia popolare – e in effetti gran parte della scienza cognitiva lo fa – come analoga ad un «diagramma a blocchi» dell'elettronica. Un riquadro del diagramma si potrebbe denominare, per esempio, «amplificatore della frequenza intermedia» pur rimanendo, questo, neutrale rispetto all'esatto circuito, e sia che l'amplificazione venga realizzata da una valvola termoionica, sia che ciò avvenga grazie ad un transistor. Rifacendoci alla terminologia impiegata da F. Jackson e P. Pettit (1988, pp. 381-400) il «*role state*» sarebbe costituito dall'«amplificatore» e il «*realiser state*» dalla «valvola termoionica». Si può quindi pensare al funzionalismo come ad una teoria della «scatola nera». Questa linea di pensiero verrà approfondita nella prossima parte.

Il fatto di ragionare in termini di relazione causale riguardo a credenze e a desideri si sposa sia con la psicologia popolare sia con le idee humane sulle ragioni dell'azione. Questo punto di vista è stato messo in discussione e criticato da alcuni filosofi; sembra tuttavia dimostrare la sua correttezza: basti pensare, ad esempio, ad un robot-aeroplano progettato per percorrere un tragitto che va da Melbourne a Sidney. Il programmatore dovrebbe inserirvi la versione elettronica di una carta geografica della regione sud-est dell'Australia. Questo elemento costituirebbe la parte delle «credenze». Bisognerebbe poi inserire un equivalente elettronico del messaggio «vai a Sidney». Ciò costituirebbe l'elemento «desiderio». Se vento o condizioni meteorologiche spingessero l'aeroplano fuori percorso, il feedback negativo riporterebbe di nuovo il velivolo sulla giusta via per Sidney. L'esistenza di meccanismi intenzionali (*purpousive*) ha finalmente dimo-
stra-

to ai filosofi (spero) che non vi è niente di misterioso nella teleologia, e che anche l'intenzionalità (con la «z»)⁴⁸ non presenta grandi problemi semantici. Per esempio, la frase «Joe desidera un unicorno» non è uguale a «Joe calcia un pallone». Perché Joe possa calciare un pallone si avrà bisogno di un pallone; nell'altro caso, non esiste nessun unicorno. Però è possibile dire «Joe desidera-vero di se stesso “possiede un unicorno”». O più semplicemente «Joe crede-vero F» o «Joe desidera-vero F», ove ad F corrisponde una frase appropriata (v. Quine 1960, pagg. 206-160). Naturalmente, se non si vuole relativizzare un linguaggio bisogna aggiungere: «o un equivalente» (*samesayer*) di F, oppure utilizzare la parola «proposizione» e questo comporta la nozione di proposizione o di intertraducibilità. Anche non volendo accettare il concetto quineiano di indeterminatezza della traduzione, ci si trova di fronte ad una certa confusione rispetto ai concetti di «credenza» e «desiderio» – confusione dovuta a quella connessa ai termini «analicità» e «sinonimia». Un teorico dell'identità potrebbe sostenere che questa confusione corrisponde alla confusione dello stato cerebrale che costituisce la credenza o il desiderio. Ma quante interconnessioni vi sono in una credenza o desiderio? Da un punto di vista olistico, come quello di Lewis, non bisogna supporre che l'individuazione di credenze e desideri sia precisa, anche se riconosciuta sufficiente sia dalla psicologia popolare che dalla metaetica di Hume. Il modo in cui il cervello rappresenta il mondo potrebbe non essere come un linguaggio, bensì come una carta geografica. Ogni elemento di una carta è collegato a tutti gli altri, anche se essa contiene un numero limitato di informazioni. Le mappe, poi, non sono costituite da un numero infinito di parti, né, tanto meno, da tante parti continue. Le credenze esprimono le diverse parti di informazione che potrebbero essere ricavate dalla carta geografica. In quest'ottica, dunque, le credenze si avvicinerrebbero molto a

⁴⁸ È lo stesso Smart che, nel testo originale, distingue tra «intentionality» con la «t» e «intensionality» con la «s». In italiano vi è situazione analoga tra «intensionalità» e «intenzionalità» (*Ndt*).

quelle individualiste, caratteristiche della psicologia popolare e della psicologia humana.

6. *Le teorie dell'identità type-token*

I concetti di «type» e «token»⁴⁹ vengono qui considerati in analogia ai «type» e «token» in ambito linguistico. Il telegramma «amore e amore e amore», ad esempio, contiene solo due *type*, ma l'impiegato telegrafico potrebbe, invece, insistere che contiene cinque parole (*token*). Analogamente, secondo la teoria dell'identità del *token*, un particolare dolore (più esattamente un sentir dolore) è identico ad un particolare processo cerebrale. Un funzionalista potrebbe essere d'accordo con questa affermazione. Di fatto il funzionalismo veniva visto come una sorta di perfezionamento della teoria dell'identità, ma anche come qualcosa di incompatibile con essa, in ragione della corretta affermazione secondo cui uno stato funzionale può divenire cosciente attraverso stati cerebrali molto diversi: così uno stato funzionale può essere compreso attraverso un cervello a base di silicio o attraverso un cervello a base di carbonio, e, tralasciando la robotica e la fantascienza, posso sostenere di percepire il mio mal di denti attraverso un processo neurale diverso da quello attraverso cui tu realizzi il tuo.

Un funzionalista può in fondo accettare le identità *token*. Generalmente, invece, i funzionalisti negano l'esistenza di identità *type*. Tuttavia Jackson, Pargetter e Prior (1982), Braddon-Mitchell e Jackson (1996) sostengono che questa è una reazione eccessiva da parte dei funzionalisti (in realtà essi vedono il funzionalismo come un ramo della teoria dell'identità). Secondo un funzionalista avere degli stati mentali equivarrebbe a dire di avere un qualsiasi stato (a base di silicio o a base di carbonio, per esempio) che spieghi le proprietà funzionali. Nel funzionalismo lo stato di secondo ordi-

⁴⁹ Si è preferito lasciare in originale i termini «type» e «token», la cui usuale traduzione in italiano con «tipo» e «occorrenza» non sembra ancora del tutto attestata a livello scientifico (*Ndt*).

ne è uno stato che possiede un certo stato di primo ordine, o un altro stato, che causa il o è causato dal comportamento al quale allude il funzionalista. In questo modo abbiamo una teoria del *type* di secondo ordine. Riflettiamo ora sulla fragilità. Sia la fragilità di un vetro che quella di un biscotto rappresentano entrambe lo stato che ha la proprietà attraverso la quale viene spiegata la loro rottura, sebbene la proprietà fisica di primo ordine sia da considerare diversa nei due casi. Questo modo di vedere la questione forse è più plausibile in relazione agli stati mentali come credenze e desideri, piuttosto che come esperienze immediatamente riferite. Quando dico di avere mal di denti mi sembra che siano coinvolte proprietà di primo ordine anche se si tratta di proprietà di natura topico-neutrale. Se continuiamo a occuparci di proprietà di primo ordine, possiamo affermare che la distinzione tra *token* e *type* è tutto o niente. Infatti, le esperienze umane sono processi cerebrali di una certa categoria e le esperienze di Alfa-Centauriani, ad esempio, sono processi cerebrali di un'altra categoria. In realtà, possiamo proporre una classificazione più accurata, senza toccare il limite dell'identità *token*.

Quanto dovrebbe essere ristretta la restrizione di una teoria ristretta del *type*? Quanti capelli al massimo deve avere un uomo pelato? Un teorico dell'identità si aspetta che il suo attuale mal di denti sia molto simile a quello che aveva ieri, e che questo, a sua volta, sia analogo a quello della moglie, e che persino, a volte, possano essere presenti delle somiglianze con quello del suo gatto. Ma non sarebbe sicuro della similarità del suo dolore col dolore di un extra-terrestre. Tuttavia anche qui troverebbe alcune somiglianze.

Quindi nel caso di analogie tra il mio attuale dolore e quello di dieci minuti fa, ci saranno alcune differenze poco importanti, così come ce ne saranno tra il mio e il tuo dolore. Pensiamo ora alla topiaria, utilizzando un'analogia che Quine ha sfruttato affrontando un'altra questione. Nei giardini della campagna inglese le siepi di bosso spesso vengono tagliate in varie forme, per esempio a forma di pavone. Si potrebbe generalizzare circa le siepi di bosso a forma di pavone e qualcun altro potrebbe affermare che tutte le imitazioni di pavone di

una particolare siepe hanno la stessa forma. Tuttavia se ci avviciniamo alle due imitazioni del pavone e le scrutiamo attentamente per notare la forma precisa dei rametti che le sostengono troveremo delle differenze. Sia che sosteniamo che due cose sono simili, sia che non lo sono, è comunque una questione di imprecisione nella descrizione. Se ci spingessimo sino al limite della precisione e della concretezza i *type* si restringerebbero a singoli *type* assimilabili esclusivamente tra loro (*single membered type*), ma ancora una volta non ci sarebbe una differenza ontologica tra teoria dell'identità e funzionalismo.

Un'interessante forma della teoria dell'identità del *token* è il monismo anomalo di Davidson (1980). Egli sostiene che le relazioni causali rientrano nelle descrizioni neurali, ma non nelle descrizioni del linguaggio della psicologia. Queste ultime utilizzano predicati intenzionali, ma, a causa dell'indeterminatezza della traduzione e dell'interpretazione, tali predicati non rientrano di diritto nella categoria delle espressioni (*law statements*). Ne segue che le identità mente-cervello possono manifestarsi solo a livello di eventi individuali (*token*). Al di là della portata del presente saggio, bisogna prendere attentamente in considerazione l'ingegnoso approccio di Davidson, poiché si differenzia in modo significativo dalle più note forme della teoria dell'identità.

7. Coscienza

Alla domanda «la coscienza è un processo cerebrale?», Place ha risposto in modo affermativo. Ma di che tipo di processo cerebrale si tratta? È naturale avvertire che esista qualcosa di indescrivibile, qualcosa che nessun mero processo neurofisiologico (caratterizzato solo da proprietà fisiche intrinseche) potrebbe avere. La sfida del teorico dell'identità è quella di dissolvere questa sensazione.

Supponiamo che io sia in sella alla mia bicicletta e che percorra il tragitto da casa mia fino all'università. Improvvisamente mi rendo conto che ho oltrepassato un ponte sul fiume, percorso un sentiero tortuoso di 800 metri, evitato il

traffico, e così via; eppure non ricordo nulla di tutto ciò. Da un lato ero cosciente: attento e informato sulla mia posizione e la mia velocità, sul percorso della mia bicicletta e sulla strada, sulle posizioni e le velocità delle macchine vicine e persino sulla lunghezza del ponte. Ma allo stesso tempo non ero cosciente: ero, per così dire, in modalità «pilota automatico». Quindi lasciatemi usare la parola «consapevolezza» (*awareness*) per descrivere questa sorta di coscienza automatica o subcosciente. Forse non ero al 100% in modalità «pilota automatico». In un certo senso potevo essere distratto e pensare alla filosofia. Questo non sarebbe stato rilevante per il mio tragitto. Ci si potrebbe domandare se si è sempre col pilota automatico; forse si spera che non sia così, specialmente rispetto all'esempio di Armstrong, relativo ad un camionista di lunga tratta (Armstrong 1962). Probabilmente è ciò che succede anche in questo caso, e se fosse realmente così il camionista sarebbe cosciente solamente perché lui, o lei, è vigile sul suo percorso, il traffico, etc., mentre è attento nel senso di «arrivare a credere attraverso i sensi». Il camionista ha delle certezze, ma non si rende conto di averle. Non c'è traccia di indescribibilità per questo senso di «coscienza», al quale mi riferirò con il termine di «consapevolezza» (*awareness*).

Per una totale coscienza, quella che ci confonde e ci fa pensare a qualcosa di indescribibile, abbiamo bisogno del significato spiegato da Armstrong in un dibattito con Norman Malcom (Armstrong e Malcom, 1962, pag. 110), considerando però che punti di vista simili sono stati espressi talvolta anche da altri filosofi, quali per esempio Savage (1976), Dennet (1991), Lycan (1996) e Rosenthal (1996). Una recente analisi di ciò può essere ritrovata in Smart (2004). Nel dibattito con Norman Malcom, Armstrong paragona la coscienza alla propriocezione (*proprioception*). Un caso di propriocezione si verifica quando, con gli occhi chiusi e senza toccare, ci rendiamo immediatamente conto dell'angolazione di uno dei nostri gomiti. La propriocezione è un senso speciale, diverso dalla sensazione corporea, attraverso cui ci rendiamo conto delle parti del nostro corpo. Il cervello è parte del nostro corpo e quindi, probabilmente, un'immediata consapevolezza di

un processo nel cervello o di uno stato cerebrale potrebbe essere considerata come una «propriocezione». Propriocezione e neuroanatomia d'altronde non coincidono. La propriocezione, che costituisce la coscienza, diversa dalla mera consapevolezza, è un ordine più alto di consapevolezza, ovvero la percezione di una parte (o configurazione) del nostro cervello attraverso il cervello stesso. Qualcuno potrebbe percepire un senso di circolarità. Se così fosse, lasciamo credere a costui che la propriocezione si verifica in un tempo trascurabile di attuazione dopo che il processo è stato propriocetto (*propriocepted*). Probabilmente esistono propriocezioni di propriocezioni, propriocezioni di propriocezioni di propriocezioni e così via, sebbene in realtà la sequenza non vada oltre il secondo o terzo stadio. L'ultima propriocezione della sequenza non sarà propriocetta (*propriocepted*), e questo potrebbe essere utile per spiegare il nostro senso di indescrivibilità della coscienza. Pensiamo al Gilbert Ryle di *The Concept of Mind on the Systematic Elusiveness of "I"* (Ryle 1949, pagg. 195-8).

Place afferma che la funzione di «pilota automatico», alla quale si riferisce definendola come «lo zombi dentro di noi», serve perché la coscienza identifichi gli input che considera problematici, e allo stesso tempo ignori gli input non problematici, o li trasformi in output senza la necessità di una consapevolezza (*awareness*) cosciente. Per questa visione della coscienza cfr. anche Place (1999).

8. Ulteriori obiezioni alla teoria dell'identità

Dovremmo qui far menzione delle autorevoli critiche della teoria dell'identità rispettivamente di Saul Kripke e di David Chalmers. Non sarà possibile analizzarle nei dettagli, in parte perché le obiezioni mosse da Kripke si basano su temi riguardanti la modalità, la semantica dei mondi possibili e l'essenzialismo, argomenti questi che alcuni filosofi vorrebbero contestare; in parte perché il libro di Chalmers, vasto e denso, meriterebbe una risposta più articolata.

Kripke (1980) definisce un'espressione un designatore rigido se si riferisce allo stesso oggetto in ogni mondo possibile.

O nella teoria del riferimento diretto (*counterpart theory*) avrebbe un esatto e simile riferimento diretto (*counterpart*) in ogni mondo possibile. A me sembra che ciò che noi consideriamo come riferimento diretto (*counterpart*) sia altamente contestuale. Prendiamo l'esempio «acqua è H₂O». In un altro mondo, o come immagina Putnam in una terra gemella nel nostro mondo (1975), la sostanza trovata nei fiumi, nei laghi o nel mare non sarebbe H₂O ma XYZ, e ciò non sarebbe acqua. Qui si sta certamente preferendo la chimica reale a quella «folk», e fino ad ora sono d'accordo. In alcuni contesti potremmo dire che nella terra gemella o nel possibile mondo immaginato la sostanza trovata nei fiumi non è acqua. Tuttavia, si possono immaginare alcuni contesti in cui il fatto che questa sostanza si trovi nei fiumi, nei laghi e nel mare, calmando la sete e sostenendo la vita, è più importante della sua composizione chimica diventando così XYZ la copia di H₂O.

Kripke riflette sull'identità «calore = movimento molecolare», sostenendo che ciò è vero in ogni mondo possibile, per cui questa identità è necessariamente vera. In realtà l'asserzione non è del tutto esatta. Che dire in effetti del calore radiante? Che dire del calore, così come viene definito dalla termodinamica classica, ovvero «topico-neutrale», rispetto alla termodinamica statistica? E ancora, supponiamo che il calore abbia un'essenza costituita da movimento molecolare, o almeno lo sia nel contesto immaginato. Kripke afferma (1980, pag. 151) che quando pensiamo che il movimento molecolare potrebbe esistere in assenza di calore noi ci stiamo confondendo, stiamo cioè pensando che il movimento molecolare potrebbe essere esistito senza essere percepito come calore. Si domanda se è analogamente possibile che se il dolore è una specie di processo cerebrale esso sia esistito senza essere percepito come dolore. Egli ritiene che la risposta sia «no». Un teorico dell'identità che accettasse il valore della coscienza come la più alta categoria di percezione potrebbe rispondere «sì». Dovremmo sapere di avere un dente guasto e anche di trovarci in una condizione di agitazione (per usare un termine di Ryle sugli stati emozionali) senza essere consci della nostra consapevolezza. Un teorico dell'identità come Smart

preferirebbe parlare di «avere un dolore» piuttosto che di «dolore»: il dolore non fa parte del contenuto del mondo tanto quanto non ne fa parte un senso *datum* o l'idraulico medio. Kripke (pag. 152) ne conclude che l'apparente contingenza della connessione tra lo stato fisico e il corrispondente stato cerebrale non può essere spiegata da una sorta di analogia qualitativa (*qualitative analogie*), come nel caso del calore.

Smart avrebbe detto che vi è una logica dietro al fatto che il collegamento di sensazioni (*sensings*) e processi cerebrali è solo in parte contingente. Una descrizione completa dello stato o processo cerebrale (includendo cause ed effetti dello stesso) implicherebbe il resoconto di esperienze interne ma, essendo quest'ultimo topico-neutrale e quindi molto astratto, non presupporrebbe la descrizione neurologica.

Durante i suoi approfonditi studi sulla coscienza, Chalmers sviluppò una teoria sui *qualia* non fisici che per certi versi evita i problemi relativi ai *nomological danglers*. La preoccupazione espressa da Smart (1959) riguarda il fatto che se esistessero *qualia* non fisici, più plausibilmente dovrebbero esserci leggi capaci di mettere in relazione i processi neurofisiologici con proprietà evidentemente semplici, e le leggi di correlazione dovrebbero essere fondamentali, meri *danglers* provenienti dalla rete nomologica (secondo la definizione della scienza di Feigl). Chalmers controbatte a questa proposta ipotizzando che i *qualia* non siano semplici, che ci sia sconosciuto il fatto che essi siano composti da semplici proto-*qualia*, e che le leggi fondamentali che collegano questi ultimi alle entità fisiche li colleghino anche alle entità fisiche fondamentali. La sua opinione richiama un panpsichismo piuttosto interessante. D'altro canto, se la spiegazione sul topico-neutrale è corretta, allora i *qualia* non sarebbero che dei punti in uno spazio di similarità multidimensionale, e l'innegabile plausibilità darebbe ragione al teorico dell'identità.

Nel caso della teoria di Chalmers, come si può essere coscienti dei *qualia* non fisici? Si è suggerito precedentemente che questa consapevolezza interna sia una propriocezione del cervello da parte del cervello. E in caso di consapevolezza di un *qualia*? Chalmers potrebbe tentare di rispondere usando il

proprio principio di coerenza. In base ad esso vi sarebbe un parallelismo tra la storia neurologica causale e la storia della successione dei *qualia*. Comunque non è chiaro se ciò ci renderebbe coscienti dei *qualia*. I *qualia* sembrerebbero non essere necessari nella storia fisiologica di come un'antilope fugge da una tigre.

Spesso la gente pensa che se anche un robot fosse in grado di scansionare i propri processi percettivi, ciò non significherebbe che il robot sia cosciente. Questo sembrerebbe facilmente intuibile, ma si potrebbe forse invertire il discorso ed affermare che siccome il robot può sapere di essere consapevole (*aware of its awareness*), allora è cosciente. Ho già menzionato in precedenza alcune ragioni per le quali diffidare delle intuizioni, ma in ogni caso Chalmers si diverte ad ipotizzare che un termostato abbia una specie di proto-*qualia*. La disputa tra teorici dell'identità (e fisicalisti in genere) e Chalmers si basa sul nostro atteggiamento rispetto alla fenomenologia. Certamente camminando in una foresta, vedendo il blu del cielo, il verde degli alberi, il rosso del sentiero, sembrerebbe difficile credere che i nostri *qualia* siano semplici punti in uno spazio di similarità multidimensionale. Ma forse questo è qualcosa che assomiglia (per usare una frase di cui si può diffidare) all'essere consapevoli di un punto in uno spazio di similarità multidimensionale. Si può anche essere soggetti, come suggerirebbe Place, a «fallacia fenomenologica». Al termine del suo libro Chalmers riflette sull'interpretazione della meccanica quantistica. Se avessero ragione, potremmo forse considerare la teoria di Chalmers come integrata nella fisica e lui, dopo tutto, come un fisicalista. Si potrebbe comunque porsi la domanda se noi dobbiamo scendere al livello dei quanti per poter capire la coscienza o se la coscienza sia rilevante per la meccanica quantistica.

Bibliografia

- ARMSTRONG, D.M., *Perception and the Physical World*, London, Routledge, 1961.
- , *Bodily Sensations*, London, Routledge, 1961.
- , *Consciousness and Causality and Reply*. In ARMSTRONG, D.M., MALCOLM, N., *Consciousness and Causality*, Oxford, Blackwell, 1962.
- , (1968), *A Materialist Theory of the Mind*, London, Routledge, 1993.
- , *The Headless Woman Illusion and the Defence of Materialism*, in «Analysis», 29, 1968, pp. 48-49.
- , *The Mind-Body Problem: An Opinionated Introduction*, Boulder, Colorado, Westview Press, 1999.
- ARMSTRONG, D.M., MARTIN, C.B., PLACE, U.T., *Dispositions: A Debate*, in T. CRANE (a cura di), London, Routledge, 1996.
- BRADDON-MITCHELL, D., JACKSON, F., *Philosophy of Mind and Cognition*, Oxford, Blackwell, 1996.
- BROAD, C.D., *The Mind and its Place in Nature*, London, Routledge and Kegan Paul, 1937.
- CAMPBELL, K., *Body and Mind*, Indiana, University of Notre Dame Press, 1984.
- CARNAP, R. (1932), 'Psychologie in Physikalischer Sprache', in «Erkenntnis», 3, pp. 107-142. Traduzione inglese in A.J. Ayer (a cura di), *Logical Positivism*, Glencoe, Illinois, Free Press 1959.
- , (1963), 'Herbert Feigl on Physicalism'. In SCHILPP, 1963, pp. 882-886.
- CHALMERS, D.M., *The Conscious Mind*, New York, Oxford University Press, 1966.
- CLARK, A., *Sensory Qualities*, Oxford, Oxford University Press, 1993.
- DAVIDSON, D., 'Mental Events', 'The Material Mind' and 'Psychology as Part of Philosophy'. In Davidson, D. *Essays on Actions and Events*, Oxford, Clarendon Press, 1980.

- DENNETT, D.C., *Consciousness Explained*, Boston, Little and Brown, 1991.
- FARRELL, B.A., 'Experience', in «Mind», 50, 1950, pp. 170-198.
- FEIGL, H. (1958), *The "Mental" and the "Physical"*. In FEIGL, H., SCRIVEN, M., MAXWELL, G. (a cura di) *Concepts, Theories and the Mind-Body Problem*, Minneapolis, Minnesota Studies in the Philosophy of Science, Vol. 2, ristampato con un Poscritto in FEIGL 1967.
- , *The 'Mental' and the 'Physical', The Essay and a Postscript*, Minneapolis, University of Minnesota Press, 1967.
- HEIL, J., *Cause, Mind and Reality, Essays Honoring C.B.Martin*, Dordrecht, Kluwer Academic Publishers, 1989.
- HILBERT, D.R., *Color and Color Perception: A Study in Anthropocentric Realism*, Stanford, CSLI, 1987.
- HILL, C.S. 1991: *Sensations: A Defense of Type Materialism*, Cambridge, Cambridge University Press.
- JACKSON, F., 'What Mary didn't know', and 'Postscript on qualia'. In JACKSON, F., *Mind, Method and Conditionals*, London, Routledge, 1998.
- JACKSON, F., PETTIT, P., 'Functionalism and Broad Content', in «Mind», 97, 1988, pp. 381-400.
- JACKSON, F., PARGETTER, R., PRIOR, E., 'Functionalism and Type-Type Identity Theories', in «Philosophical Studies», 42, 1982, pp. 209-225.
- KIRK, R., 'Why There Couldn't be Zombies', in «Aristotelian Society», Supp. Vol. 73, 1999, pp. 1-16.
- KRIPKE, S., *Naming and Necessity*, Cambridge, Mass., Harvard University Press, 1980.
- LEVIN, M.E., *Metaphysics and the Mind-Body Problem*, Oxford, Clarendon Press, 1979.
- LEWIS, D., 'An Argument for the Identity Theory', in «Journal of Philosophy», 63, 1966, pp. 17-25.
- , 'How to Define Theoretical Terms', in «Journal of Philosophy», 67, 1970, pp. 427-446.
- , 'Psychophysical and Theoretical Identifications', in «Australasian Journal of Philosophy», 50, 1972, pp. 249-258.

- , *'Mad Pain and Martian Pain' and 'Postscript'*. In LEWIS D. «Philosophical Papers», Vol. 1, Oxford, Oxford University Press, 1983.
- , *'What Experience Teaches'*. In LYCAN, W. (a cura di), *Mind and Cognition*, Oxford, Blackwell, 1989.
- , *'Reduction of Mind'*. In GUTTENPLAN, S. (a cura di), *A Companion to the Philosophy of Mind*, Oxford, Blackwell, 1994.
- LYCAN, W.G., *Consciousness and Experience*, Cambridge, Mass., M.I.T. Press, 1996.
- MEDLIN, B.H., *'Ryle and the Mechanical Hypothesis'*. In Presley, C.F. (a cura di), *The Identity Theory of Mind*, St. Lucia, Queensland, Queensland University Press, 1967.
- , *'Materialism and the Argument from Distinct Existences'*. In MACINTOSH, J.J., COVAL, S., *The Business of Reason*, London, Routledge and Kegan Paul, 1969.
- PITCHER, A., *Theory of Perception*, Princeton, N.J., Princeton University Press, 1971.
- PLACE, U.T., *'The Concept of Heed'*, in «British Journal of Psychology», 45, 1954, pp. 243-255.
- , *'Is Consciousness a Brain Process?'*, in «British Journal of Psychology», 47, 1956, pp. 44-50,
- , *'Materialism as a Scientific Hypothesis'*, in «Philosophical Review», 69, 1960, pp. 101-104.
- , *'Comments on Putnam's "Psychological Predicates"'*. In CAPITAN, W.H., MERRILL, D.D. (a cura di) *Art, Mind and Religion*, Pittsburgh, Pittsburgh University Press, 1967.
- , 1988: *'Thirty Years on--Is Consciousness still a Brain Process?'*, in «Australasian Journal of Philosophy», 66, 1988, pp. 208-219.
- , *'Low Claim Assertions'*. In HEIL, J. (a cura di), *Cause, Mind and Reality: Essays Honoring C.B. Martin*, Dordrecht, Kluwer Academic Publishers, 1989.
- , *'E.G. Boring and the Mind-Brain Identity Theory'*, in «British Psychological Society, History and Philosophy of Science Newsletter», 11, 1990, pp. 20-31.
- , *'Connectionism and the Problem of Consciousness'*, in «Acta Analytica», 22, 1999, pp. 197-226.

- , *Identifying the Mind*, New York, Oxford University Press, 2004.
- PUTNAM, H., ‘*Minds and Machines*’. In HOOK, S. (a cura di), *Dimensions of Mind*, New York, New York University Press, 1960.
- , *The Meaning of “Meaning”*. In PUTNAM, H. *Mind, Language and Reality*, Cambridge, Cambridge University Press, 1975.
- QUINE, W.V., *Word and Object*, Cambridge, Mass., MIT Press, 1960.
- REICHENBACH, H., *Experience and Prediction*, Chicago, Chicago University, 1938
- ROSENTHAL, D.M., ‘*Identity Theories*’. In GUTTENPLAN, S. (a cura di), *A Companion to the Philosophy of Mind*, Oxford, Blackwell, 1994, pp. 348-355.
- , ‘*A Theory of Consciousness*’. In BLOCK, N., FLANAGAN, O. GÜZELDERE, G. (a cura di) *The Nature of Consciousness*, Cambridge, Mass., MIT Press, 1966.
- RYLE, G., *The Concept of Mind*, London, Hutchinson, 1949.
- SAVAGE, C.W., ‘*An Old Ghost in a New Body*’. In GLOBUS, G.G., MAXWELL, G., SAVODNIK, I., (a cura di), *Consciousness and the Brain*, New York, Plenum Press, 1976.
- SCHILPP, P.A. (a cura di), *The Philosophy of Rudolf Carnap*, La Salle, Illinois, Open Court, 1963.
- SCHLICK, M., ‘*De la Relation des Notions Psychologiques et des Notions Physiques*’, in «*Revue de Synthèse*», 10, pp. 5-26, 1935. Traduzione inglese in in FEIGL, H., SELLARS, W., *Readings in Philosophical Analysis*, New York, Appleton-Century Crofts, 1949.
- SMART, J.J.C., ‘*Sensations and Brain Processes*’, in «*Philosophical Review*», 68, 1959, pp. 141-156.
- , ‘*Colours*’, in «*Philosophy*», 36, 1961, pp. 128-142.
- , ‘*Materialism*’, in «*Journal of Philosophy*», 60, 1963, pp. 651-662.
- , ‘*On Some Criticisms of a Physicalist Theory of Colour*’. In CHUNG-YING CHENG (a cura di), *Philosophical Aspects of the Mind-Body Problem*, Honolulu, University of Hawaii Press, 1975.

La teoria dell'identità della mente

- , *'The Content of Physicalism'*, in «Philosophical Quarterly», 28, 1978, pp. 339-341.
- , *'Physicalism and Emergence'*, in «Neuroscience», 6, 1981, pp. 109-113.
- , *"Looks Red" and Dangerous Talk*, in «Philosophy», 70, 1995, pp. 545-554.
- , *'Consciousness and Awareness'*, in «Journal of Consciousness Studies», 11, 2004, pp. 41-50.

LA RAPPRESENTAZIONE MENTALE⁵⁰
Definizioni, modelli, teorie

David Pitt

Una «rappresentazione mentale» è anzitutto un concetto teoretico della scienza cognitiva e, in quanto tale, costituisce una delle idee di base della teoria computazionale della mente, secondo la quale gli stati ed i processi cognitivi sono dati dalla verifica, dall'elaborazione e dall'immagazzinamento (nella mente o cervello) di strutture (rappresentazioni) che contengono informazioni di vario genere.

Supponendo, tuttavia, che una rappresentazione sia un oggetto dotato di proprietà semantiche (contenuto, riferimenti, condizioni-verità, valore-verità, ecc.), una rappresentazione mentale potrebbe essere considerata, in linea di massima, come un oggetto mentale dotato di proprietà semantiche. Pertanto, la rappresentazione mentale (stati e processi compresi) non deve essere intesa solo da un punto di vista computazionale. Sulla base di questa interpretazione più ampia, una rappresentazione mentale è un concetto filosofico che affonda le radici nell'antichità e che vanta una letteratura ed una tradizione fiorenti, che precedono di gran lunga la recente «rivoluzione cognitiva». Nonostante la maggior parte dei filosofi della mente contemporanei riconosca l'importanza della scienza cognitiva, i metodi ed i risultati di tale scienza è spesso difforme. Per molti di loro vanno ancora affrontati i problemi relativi alle proprietà rappresentazionali della men-

⁵⁰ D. Pitt, *Mental Representation*, in «Stanford Encyclopedia of Philosophy», <http://plato.stanford.edu/entries/mental-representation/>

te, che possono essere affrontati indipendentemente dall'ipotesi computazionale.

Sebbene i termini «teoria rappresentazionale delle mente» e «teoria computazionale della mente» vengano a volte usati come sinonimi, in questa sede mi riferirò ad essa per indicare qualsiasi teoria che presupponga l'esistenza di oggetti mentali dotati di valore semantico, inclusi i capisaldi filosofici della mente (pensieri, concetti, percetti, idee, impressioni, regole, schemi, immagini, illusioni, ecc.), così come i diversi tipi di rappresentazioni «subpersonali» postulati dalla scienza cognitiva. Le teorie rappresentazionali possono pertanto entrare in conflitto con quelle teorie che negano l'esistenza di tali oggetti: mi riferisco in particolare alle teorie di Baker (1995), Collins (1987), Dennet (1987), Gibson (1966, 1979), Reid (1764/1997), Stich (1983) e Thau (2002).

1. *La teoria rappresentazionale della mente*

La teoria rappresentazionale della mente (TRM), che trova la sua prima elaborazione nell'opera di Aristotele, è stata sviluppata a partire dagli stati mentali propri della psicologia del senso comune, quali pensieri, credenze, desideri, percezioni ed immagini. Si ritiene che questi stati abbiano una certa «intenzionalità»: riguardano o si riferiscono a qualcosa di specifico, e potrebbero essere valutati sulla base di proprietà quali coerenza, verità, appropriatezza, esattezza. Ad esempio, il pensiero secondo cui i cugini non sono parenti è priva di coerenza, la convinzione che Elvis sia morto è vera, il desiderio di mangiare la Luna è inappropriato; analogamente, l'esperienza visiva di una fragola rossa è dotata di esattezza, almeno quanto ne è priva quella di George W. Bush con le trecce rasta.

La TRM afferma che tali stati mentali discendono dalle rappresentazioni mentali: la loro già citata intenzionalità si spiegherebbe facendo riferimento alle proprietà semantiche delle rappresentazioni stesse. Ad esempio, la convinzione che Elvis sia morto deve essere considerata come discendente dalla rappresentazione mentale il cui contenuto preposizionale è

«Elvis è morto». Allo stesso modo, desiderare la morte di Elvis, temerla, oppure dispiacersene, sono tutti discendenti dalla stessa rappresentazione mentale, con la quale avranno stabilito diverse forme di relazione. Percepire una fragola, allo stesso modo, significa avere una specifica esperienza sensoriale che discende dalla fragola stessa (la fragola ne può essere la causa, per esempio).

La TRM, inoltre, definisce processi mentali (ad esempio pensare, ragionare o immaginare) come sequenze di stati mentali intenzionali. Ad esempio, immaginare la Luna che sorge dietro una montagna significa richiamare una serie di immagini mentali, tipiche della luna e di una montagna. Dedurre una proposizione q da un'altra proposizione p («se p allora q ») comporta, tra le altre cose, anche la sequenza di pensieri che ricalca la formula p , se p allora q , q .

Una delle convinzioni che accomuna i filosofi della mente contemporanei è la supposizione (o meglio, la speranza) che la mente possa essere naturalizzata (vale a dire, che a tutti gli eventi mentali si possa trovare una spiegazione in termini di scienze naturali). Questa ipotesi è condivisa dalla scienza cognitiva, per la quale l'origine degli stati e dei processi mentali sarebbe da ricondursi, in ultima analisi, alle caratteristiche fisiche del cervello e del sistema nervoso centrale. Movendo da questo stesso presupposto, le varie sotto-discipline della scienza cognitiva (incluse la psicologia e le neuroscienze computazionali e cognitive) postulano l'esistenza di processi e strutture differenti, molti dei quali non trovano posto nell'interpretazione degli stati e dei processi mentali come emerge dalla psicologia del senso comune. In ogni caso, persiste la convinzione generale secondo cui questi ultimi debbano essere spiegati in termini di rappresentazioni mentali.

In filosofia, le dispute più recenti, relative alle rappresentazioni mentali, hanno avuto come oggetto l'esistenza di atteggiamenti proposizionali (credenze, desideri eccetera) e la determinazione del loro contenuto (in che modo arrivano a significare ciò che significano), nonché l'esistenza di proprietà fenomeniche e il loro rapporto con il contenuto dei pensieri e le esperienze percettive. All'interno della scienza cognitiva

vera e propria, invece, le dispute filosoficamente più rilevanti hanno interessato l'architettura computazionale del cervello e del sistema nervoso centrale, oltre all'eventuale compatibilità tra l'interpretazione del concetto di mentalità, relativo alla scienza da una parte e alla psicologia del senso comune dall'altra.

2. *Atteggiamenti proposizionali*

Alcuni esponenti del realismo intenzionale, come Dretske (1988) e Fodor (1987), hanno messo in evidenza come le generalizzazioni che mettiamo in atto nella nostra quotidianità per prevedere o spiegare il comportamento proprio e altrui (e che fanno capo alla cosiddetta *folk psychology*, «psicologia popolare») non solo si caratterizzano per il loro elevato margine di successo, ma possono anche essere considerate come imprescindibili. Infatti, ciò in cui una persona crede, ciò che teme o desidera, ciò di cui tende a dubitare, tutto questo non può che essere un indicatore estremamente affidabile di quale sarà il suo comportamento in futuro; inoltre, se non lo imputassimo a tali stati, applicando poi le opportune generalizzazioni, il comportamento altrui ci sembrerebbe senza dubbio incomprensibile. Pertanto, dobbiamo dare atto della veridicità fondamentale delle conclusioni tratte dalla psicologia del senso comune e, quindi, dell'esistenza di quegli stati cui le sue generalizzazioni si riferiscono.

Gli eliminativisti intenzionali, quali Churchland, Dennett (probabilmente) e Stich (per un certo periodo di tempo), ritengono che gli atteggiamenti proposizionali, così come i loro stati rappresentazionali costituenti, non siano fondamentali nella corretta spiegazione e previsione dei nostri comportamenti. Churchland rigetta, in altre parole, la veridicità di quella psicologia del senso comune che fa leva sugli atteggiamenti proposizionali. Churchland (1981) crede inoltre che la psicologia ingenua sia una teoria della mente con alle spalle una lunga tradizione di fallimenti, che però, nonostante tale declino, ancora resiste ad una incorporazione nel quadro delle moderne teorie scientifiche (la psicologia cognitiva, per e-

sempio). La potremmo mettere, in altre parole, sullo stesso piano dell'alchimia, o delle teorie flogistiche e con ogni probabilità un destino altrettanto grigio la starebbe attendendo. La psicologia del senso comune è menzognera: gli stati e le rappresentazioni da essa postulati, molto semplicemente, non esistono. (Si noti comunque il fatto che quello di Churchland non è un eliminativismo *tout court*: non investe, ad esempio, il concetto intero di rappresentazione mentale; cfr. Churchland 1989).

Dennett (1987a), invece, assicura che le generalizzazioni della psicologia del senso comune siano vere ed indispensabili; tuttavia, per lui, questa non è affatto una ragione sufficiente per credere alle entità cui tale psicologia sembra riferirsi. Sostiene, dunque, che per dare una spiegazione intenzionale del comportamento di un sistema occorre, semplicemente, assumere una «posizione intenzionale» verso tale comportamento. Assumiamo come vera l'ipotesi secondo cui assegnare stati dotati di contenuto ad un sistema, e pertanto prevederne e spiegarne il comportamento, sia una strategia di successo. Tale ipotesi comporta, in altre parole, credere che tale sistema sia razionale, ad esempio che, dati gli atteggiamenti proposizionali che dovrebbe avere in un dato ambiente, il sistema si comporti come dovrebbe. Partendo dal presupposto che tale ipotesi sia vera, allora il sistema ha natura intenzionale, e le generalizzazioni relative agli atteggiamenti proposizionali sono altrettanto vere (cfr. Dennett 1987a: 29). (Anche Davidson 1973, 1974 e Lewis 1974 erano dell'opinione che ciò che possiede un atteggiamento proposizionale deve essere interpretabile solo in un modo determinato. Comunque, non è del tutto chiaro se essi ritenessero irreali gli atteggiamenti proposizionali).

Stich (1983) sostiene che tra gli obiettivi della psicologia cognitiva non ci sia (e, in ogni caso, non dovrebbe mai esserci) la classificazione quasi tassonomica degli stati mentali sulla base delle loro proprietà semantiche. La classificazione degli stati mentali sulla base del loro contenuto dipende, infatti, da una serie di fattori che la rendono particolarmente problematica nel contesto di una psicologia scientifica. La psico-

logia cognitiva cerca le spiegazioni causali del comportamento e della cognizione, ed afferma che il potere causale di uno stato mentale è determinato dalle sue proprietà intrinseche «strutturali» o «sintattiche». Tuttavia, le proprietà semantiche di uno stato mentale sono determinate dalle sue proprietà estrinseche, siano esse la sua storia, o i suoi rapporti intramentali oppure ambientali. Pertanto, tali proprietà non possono figurare tra le spiegazioni scientifico-causali del comportamento. (Fodor 1994 e Dretske 1988 tentano proprio di far fronte ad alcune di queste problematiche). Stich invece propone una teoria sintattica della mente nella quale le proprietà semantiche delle rappresentazioni mentali non svolgono alcun ruolo esplicativo. (Stich ha più volte riveduto le sue posizioni a questo proposito; cfr. Stich 1996).

3. *Rappresentazione concettuale e non-concettuale*

Secondo la dottrina comunemente accettata dai realisti, esistono due tipi principali di stati rappresentazionali (cfr. Boghossian, 1995). Da un lato vi sono quegli stati, come ad esempio i pensieri, costituiti da concetti privi di caratteristiche (i cosiddetti *qualia*) fenomeniche («ciò-che-appare-come»); dall'altro lato vi sono quelli, come le esperienze sensoriali, dotati di caratteristiche fenomeniche ma privi di elementi concettuali. Il contenuto non concettuale è generalmente definito come un tipo di contenuto che individua anche gli stati di una creatura priva di concetti. Sulla base di questa tassonomia, gli stati mentali possono rappresentare, in maniera analoga, sia le espressioni dei linguaggi naturali, sia disegni, dipinti, cartine geografiche o fotografie. (Gli stati percettivi, come ad esempio vedere che un dato oggetto è blu, sono a volte considerati stati ibridi, costituiti, ad esempio, da un'esperienza sensoriale non concettuale ed un pensiero, o da un insieme più integrato di componenti sensoriali e concettuali).

Alcune analisi storiche delle proprietà rappresentazionali della mente (per esempio, Aristotele 1984, Locke 1689/1975, Hume 1739/1978) sembrano supporre che le rappresentazioni

non concettuali – oggetti di percezione («impressioni»), immagini («idee») e così via – siano il solo tipo possibile di rappresentazioni mentali e che la mente rappresenti gli oggetti del mondo che ci circonda in stati. Avallando questa ipotesi, il contenuto di tutti gli stati rappresentazionali dipende dalle relative caratteristiche fenomeniche. Tuttavia, valide argomentazioni sulla mancanza di generalità (Berkeley, 1975), sull'ambiguità (Wittgenstein, 1953) e sulla non-composizionalità (Fodor, 1981c) delle rappresentazioni sensoriali e immaginistiche, oltre che sulla loro inadeguatezza a fungere come concetti logici (Frege, 1918/1997; Geach, 1957) o matematici (Frege, 1884/1953) e sulla simmetria della somiglianza (Goodman, 1976), hanno portato i filosofi a credere che non può esistere nessuna teoria della mente che preveda solo delle rappresentazioni non-concettuali costruite in questo modo.

Le divergenze attuali sulle rappresentazioni non-concettuali riguardano l'esistenza e la natura delle proprietà fenomeniche ed il ruolo che queste hanno nel determinare il contenuto dell'esperienza sensoriale. Dennet (1988), ad esempio, nega l'esistenza stessa dei *qualia*; Brandom (2002), McDowell (1994), Rey (1991) e Sellars (1956), dal canto loro, negano che i *qualia* siano necessari per spiegare il contenuto dell'esperienza sensoriale. Tra coloro che riconoscono il contenuto fenomenico di tali esperienze, alcuni (Dretske, Lycan, Tye) sostengono che sia riducibile ad un tipo di contenuto intenzionale, mentre altri (Block, Loar, Peacocke) lo considerano non riducibile (vedi paragrafo successivo).

È controversa anche la tesi tradizionale secondo cui le rappresentazioni concettuali (pensieri, credenze) siano prive di fenomenologia. Chalmers (1996), Flanagan (1992), Goldman (1993), Horgan e Tiensen (2003), Jackendoff (1987), Levine (1993, 1995, 2001), McGinn (1991a), Pitt (2004), Searle (1992), Siewert (1998) e Strawson (1994) sostengono che gli stessi stati rappresentazionali puramente simbolici (consci) sono dotati di una fenomenologia (probabilmente proprietaria). Se questa ipotesi è corretta, la questione del ruolo della fenomenologia nella determinazione del contenu-

to si ripresenta per la rappresentazione concettuale; e le ambizioni eliminativiste di Sellars, Brandom, Rey, ecc. incontrerebbero un nuovo ostacolo. (Porrebbe inoltre dei problemi per il rappresentazionalismo riduttivista).

4. *Rappresentazionalismo e fenomenalismo*

A seconda del loro approccio alle proprietà fenomeniche, i realisti si dividono tra rappresentazionalisti (chiamati anche «rappresentazionisti» e «intenzionalisti»), tra cui, ad esempio, Dretske (1995), Harman (1990), Leeds (1993), Lycan (1987, 1996), Rey (1991), Thau (2002), Tye (1995, 2000), e fenomenalisti (detti anche «fenomenisti» e «fanatici dei *qualia*» [*qualia freaks*]), tra i quali spiccano Block (1996, 2003), Chalmers (1996 e successivi), Evans (1982), Loar (2003a, 2003b), Peacocke (1983, 1989, 1992, 2001), Raffman (1995) e Shoemaker (1990). I rappresentazionalisti sostengono che il carattere fenomenico di uno stato mentale è riducibile ad un tipo di contenuto intenzionale. I fenomenalisti, al contrario, sostengono che il carattere fenomenico di uno stato mentale non è riducibile ad alcun tipo di contenuto intenzionale.

Spesso la tesi rappresentazionalista si basa sull'idea secondo la quale le proprietà fenomeniche siano rappresentazionali o intenzionali. Tale formulazione, tuttavia, si colloca in una posizione ambigua, a metà tra un'ipotesi riduttiva ed una non riduttiva (sebbene il termine «rappresentazionalismo» sia usato più di frequente per indicare l'ipotesi riduttivista) (vedi Chalmers). Da un lato, tale formulazione potrebbe indicare che il contenuto fenomenico di un'esperienza è un tipo di contenuto intenzionale (le proprietà che rappresenta). Dall'altro lato, potrebbe indicare che sono le proprietà fenomeniche (irriducibili) di un'esperienza a determinare il contenuto intenzionale. I rappresentazionalisti, come Dretske, Lycan e Tye, concordano con la prima ipotesi, mentre i fenomenalisti, come Block, Chalmers, Loar e Peacocke, concordano con la seconda. (I fenomenalisti hanno opinioni divergenti anche sulla possibilità che i *qualia* siano intrinsecamente rappresentazionali (Loar) o meno (Block, Peacocke)).

La maggior parte dei rappresentazionalisti (riduttivisti) è motivata dalla convinzione che ogni spiegazione naturalistica d'intenzionalità (vedi paragrafo successivo) sia a grandi linee corretta, e dal desiderio di completare la naturalizzazione del mentale applicando tali teorie al problema della fenomenalità. (Non occorre spiegare come la maggior parte dei fenomenalisti – Chalmers ne costituisce l'eccezione principale – desiderino naturalizzare il fenomenico, sebbene non nello stesso modo).

L'argomentazione principale del rappresentazionalismo riguarda la trasparenza dell'esperienza (cfr Tye, 2000: 45-51). Le proprietà che caratterizzano il 'cosa si prova' ad avere un'esperienza percettuale si presentano nell'esperienza come proprietà di oggetti percepiti: nel corso di un'esperienza, ci sembra di «vedere attraverso» l'esperienza stessa, riuscendo così a vedere gli oggetti e le relative proprietà sperimentate. Questi non si presentano, dunque, come proprietà dell'esperienza stessa; se, tuttavia, fossero proprietà dell'esperienza, la percezione sarebbe decisamente ingannevole. Secondo i rappresentazionalisti, il carattere fenomenico di un'esperienza è dovuto al suo obiettivo rappresentativo, alle sue proprietà non empiriche. (Nella percezione veridica queste proprietà sono localmente istanziate, mentre nell'illusione e nell'allucinazione non lo sono). Avallando questa ipotesi, l'introspezione diventa una percezione indiretta: conosciamo le caratteristiche fenomeniche delle nostre esperienze nel momento in cui ci rendiamo conto delle caratteristiche oggettive rappresentate dalle esperienze stesse. differenze intuitive tra le rappresentazioni concettuali e quelle sensoriali, i rappresentazionalisti ricorrono alle loro differenze strutturali o funzionali. Dretske (1995), per esempio, distingue le esperienze ed i pensieri sulla base dell'origine e della natura delle loro funzioni: un'esperienza di una proprietà P è uno stato di un sistema la cui funzione evoluta è quella di indicare la presenza di P nell'ambiente; un pensiero che rappresenti la proprietà P, d'altra parte, è uno stato di un sistema la cui funzione assegnata (acquisita) è quella di calibrare la risposta di un sistema

empirico. Rey (1991) considera sia i pensieri sia le esperienze collegati alle frasi nel linguaggio del pensiero, e li distingue sulla base (dei ruoli funzionali) dei predicati costitutivi delle frasi in questione. Lycan (1987, 1996) li distingue secondo i loro profili funzional-computazionali. Tye (2000) li distingue secondo i loro ruoli funzionali e la struttura intrinseca dei loro veicoli: i pensieri sono delle rappresentazioni in una lingua come mezzo, mentre le esperienze sono delle immagini come rappresentazioni che consistono in «file di simboli» (cfr. il contenuto delle immagini mentali, Tye 2001).

I fenomenalisti tendono a servirsi degli stessi tipi di caratteristiche (funzione, struttura intrinseca) per spiegare alcune delle differenze intuitive tra i pensieri e le esperienze; tuttavia, non ritengono che tali caratteristiche siano sufficienti a determinare le differenze tra le rappresentazioni fenomeniche e quelle non-fenomeniche. Secondo i fenomenalisti, sono le proprietà fenomeniche delle esperienze (gli stessi *qualia*) a costituire la differenza fondamentale tra l'esperienza ed il pensiero. Peacocke (1992), ad esempio, sviluppa la nozione di uno «scenario» percettuale (un trasferimento di proprietà fenomeniche alle coordinate di uno spazio egocentrico tridimensionale), il cui contenuto è «corretto» (una proprietà semantica) se nella «scena» corrispondente (la parte di mondo rappresentato dallo scenario) le proprietà sono distribuite allo stesso modo che le analoghe proprietà fenomeniche nello scenario.

Un altro tipo di rappresentazione definita dai fenomenalisti (per esempio Block, Chalmers (2003) e Loar (1996)) è il «concetto fenomenico», un ibrido concettuale-fenomenico che consiste di un «campione» fenomenologico (un'immagine o una sensazione) integrato con (o funzionante come) un componente concettuale. I concetti fenomenici sono necessari per giustificare, tra gli altri, il fatto apparente che «non puoi formare concetti (introspezzivi) di proprietà coscienti a meno che tu stesso non istanzi tali proprietà» (McGinn, 1991b). Non è possibile possedere un concetto fenomenico di una proprietà fenomenica P e, dunque, credenze fenomeniche riguardo a P, senza avere esperienza di P, perché P stessa

è (in qualche modo) costitutiva del concetto di P (cfr. Jackson, 1982, 1986; Nagel, 1974).

5. *Imagery: le immagini mentali*

Sebbene le immagini mentali abbiano avuto un ruolo fondamentale nella storia della filosofia della mente, la loro importanza nella letteratura contemporanea va ricercata anzitutto nella psicologia. Una serie di esperimenti effettuati negli anni Settanta (riassunti da Kosslyn nel 1980 e da Shepard e Cooper nel 1982) hanno evidenziato come i tempi di risposta dei soggetti a cui veniva chiesto di esaminare alcune figure variassero secondo le proprietà spaziali (dimensione, orientamento, ecc.) di tali figure. Il tentativo di spiegare questi risultati sperimentali ha suscitato un vivace dibattito sulla natura delle immagini mentali e dell'immaginazione.

Secondo Kosslyn (1980), tali risultati suggeriscono che l'analisi delle figure è stata eseguita attraverso l'impiego e l'elaborazione di rappresentazioni mentali, esse stesse dotate di proprietà spaziali (come, ad esempio, le rappresentazioni figurative o le immagini). Altri, in modo particolare Pylyshyn (1979, 1981a, 1981b, 2003), sostengono che i fatti empirici possono essere spiegati solo alla luce delle rappresentazioni proposizionali e dei processi cognitivi da queste definiti. (Pylyshyn considera tali rappresentazioni come frasi che costituiscono una lingua del pensiero).

I sostenitori della visione figurativa delle immagini mentali (vedi, per esempio, Kosslyn e Pomerantz, 1977) non ritengono che le rappresentazioni figurative siano effettivamente delle figure nella testa. Essi sostengono, piuttosto, che le immagini mentali siano in grado di rappresentare un oggetto in un modo molto simile a quello delle figure. (Un'attenzione particolare è stata rivolta anche alle immagini mentali visive: da qui l'aggettivo «figurative», sebbene ovviamente le immagini mentali possano essere anche, ad esempio, uditive, olfattive, ecc.).

È possibile operare una distinzione tra le rappresentazioni figurative e quelle proposizionali dalla distinzione esistente

tra le rappresentazioni analogiche e quelle digitali. Sebbene tale distinzione sia stata interpretata in vari modi, una delle interpretazioni più comunemente accettate è quella che considera la rappresentazione analogica continua (date le proprietà continuamente variabili della rappresentazione) e quella digitale discontinua/discreta (date le proprietà di cui la rappresentazione può essere dotata o meno) (Dretske, 1981). (Si potrebbe operare una distinzione analogico/digitale anche sulla base dei processi cognitivi. Cfr. Block, 1983). Sulla base di questa distinzione analogico/digitale, le rappresentazioni di immagini (*imagistic*), le cui proprietà possono variare continuamente (possono essere, ad esempio, più o meno visibili, rumorose, vivide, ecc.), sarebbero analogiche, mentre le rappresentazioni concettuali, le cui proprietà non variano continuamente (un pensiero su Elvis non può aumentare o diminuire: esiste oppure no), sarebbero digitali.

Si potrebbe argomentare che la distinzione figurativo/proposizionale potrebbe essere meglio definita impiegando la distinzione fenomenico/non-fenomenico, ma in questo caso non è così. Anzitutto, alcune proprietà non-fenomeniche delle rappresentazioni potrebbero variare continuamente. Inoltre, alcuni modi di interpretare le rappresentazioni figurative non presuppongono né la fenomenalità né l'analogicità. Secondo Kosslyn (1980, 1982, 1983), una rappresentazione mentale è «simil-figurativa» (*quasi-pictorial*) quando ogni parte della rappresentazione corrisponde ad una parte dell'oggetto rappresentato e tra le parti di tale oggetto e quelle della rappresentazione è mantenuta la stessa distanza. Le distanze tra le parti della rappresentazione, tuttavia, possono essere definite considerando la loro funzione, piuttosto che lo spazio, ad esempio, sulla base del numero delle fasi computazionali discrete necessarie per mettere insieme le relative informazioni immagazzinate.

Secondo Tye (1991), le immagini sono rappresentazioni ibride, formate sia da elementi discorsivi che da elementi figurativi; sono «file di simboli (descritti) interpretati»: i simboli sono elementi proposizionali, mentre la loro disposizione in file ha un significato rappresentazionale (la posizione di ogni

«cella» in una fila rappresenta uno spazio bidimensionale ben definito sulla superficie dell'oggetto immaginato).

6. *Determinazione del contenuto*

Generalmente si ritiene che i contenuti delle rappresentazioni mentali siano oggetti astratti (proprietà, relazioni, proposizioni eccetera). Una domanda cui i naturalisti in particolare hanno cercato di dare risposta è, però, in che modo le rappresentazioni mentali arrivino a possedere il loro specifico contenuto. La questione è, in questo caso, non tanto in che maniera naturalizzare il contenuto (un oggetto astratto non può essere naturalizzato), quanto piuttosto in che maniera fornire una spiegazione naturalistica dei rapporti che determinano il contenuto e che intercorrono tra le rappresentazioni mentali e gli oggetti astratti da esse rappresentati. Le teorie naturalistiche contemporanee riguardanti proprio la determinazione del contenuto sono di due tipi fondamentali: quella informazionale-causale e quella funzionale.

Le teorie informazionale-causali (Dretske, 1981, 1988, 1995) affermano che il contenuto di una rappresentazione mentale sia riconducibile alle informazioni, contenute in essa, che spiegano in quali condizioni essa si verifica (Devitt, 1996) o si verificherebbe (Fodor, 1987, 1990a). In ogni caso, è generalmente condivisa l'affermazione secondo cui le relazioni informazionale-causali non siano sufficienti per determinare il contenuto delle rappresentazioni mentali. Tali relazioni sono comuni, ma la rappresentazione non lo è. Tre camion, fumo, termostati, telefoni che suonano: ciascuno è portatore di informazioni concernenti ciò da cui discendono per mezzo di un rapporto di causalità, ma non rappresentano (in senso rilevante) quel qualcosa di cui sono portatori di informazioni. Per di più, una rappresentazione può essere causata da qualcosa che non rappresenta e, viceversa, può rappresentare qualcosa che non ne è la causa.

I tentativi principali di spiegare cosa rende uno stato informazionale-causale una rappresentazione mentale sono stati la «teoria della dipendenza asimmetrica» (Fodor, 1990b,

1990a, 1994) e le «teorie teleologiche» (Fodor, 1990b; Millikan, 1984; Papineau, 1987; Dretske, 1988, 1995). La teoria della dipendenza asimmetrica distingue i rapporti informativi da quelli rappresentazionali sulla base di un criterio «gerarchico»: i rapporti informativi dipendono da quelli rappresentazionali e non viceversa. Ad esempio, se i segni di una rappresentazione mentale sono ragionevolmente causati da cavalli, tori-di-notte, zebre-nella-nebbia, eccetera, allora porteranno informazioni relative ai cavalli, ai tori-di-notte e così via. Tuttavia, se tali segni sono causati da tori-di-notte perché sono stati a loro volta causati da cavalli, ma non viceversa, allora tali segni rappresentano cavalli.

Secondo le teorie teleologiche, i rapporti rappresentazionali sono quei rapporti in cui un meccanismo di «produzione di rappresentazioni» selezionato ha assunto (per evoluzione o per apprendimento) una funzione preponderante. In altre parole, la rappresentazione di un cavallo causata, però, da una zebra non significa «zebra» perché, nel meccanismo selezionato attraverso cui tali segni sono stati prodotti, tale rappresentazione ha comunque la funzione di indicare i cavalli e non le zebre. Il meccanismo che produce la rappresentazione «cavallo» in risposta allo stimolo «zebra» è un meccanismo mal funzionante.

Secondo le teorie funzionali (Block, 1986; Barman, 1973), il contenuto di una rappresentazione mentale risiede nei rapporti (causali, computazionali, inferenziali) intrattenuti con le altre rappresentazioni mentali. Non sono però concordi sulla possibilità di includere tra gli elementi coinvolti in tali rapporti tutte le altre rappresentazioni mentali o solo alcune, né sull'eventualità di includere anche gli stati o le vicende esterne. È riconducibile all'olismo l'ipotesi secondo cui il contenuto di una rappresentazione mentale è determinato dai suoi rapporti inferenziali-computazionali con tutte le altre rappresentazioni; si parla invece di localismo (o molecolarismo) quando si ritiene che il contenuto di una certa rappresentazione sia determinato dai rapporti che essa intrattiene soltanto con alcuni tra gli altri stati mentali. (L'affermazione secondo cui il contenuto di uno stato mentale è assolutamente in-

dipendente dai rapporti stabiliti con gli altri prende il nome di atomismo). Le teorie funzionali che non riconoscono l'esistenza di alcun rapporto esterno determinante del contenuto sono le cosiddette teorie solipsistiche (Barman, 1987). Alcuni autori fanno una distinzione di ruoli tra connessioni interne ed esterne: quelle interne determinerebbero proprietà semantiche analoghe al senso, quelle esterne invece determinerebbero proprietà analoghe al riferimento (McGinn, 1982; Sterelny, 1989).

I rappresentazionalisti riduttivi (Dretske, Lycan, Tye) si avvalgono generalmente dell'una o dell'altra teoria per fornire una spiegazione del contenuto (non concettuale) degli stati empirici. Tendono pertanto ad avere un approccio externalistico (cfr. la prossima sezione) relativamente al contenuto sia fenomenologico che concettuale. D'altro canto, i fenomenalisti ed i rappresentazionalisti non riduttivi (Block, Chalmers, Loar, Peacocke, Siewert) ritengono che il contenuto rappresentazionale di tali stati sia (almeno in parte) determinato dalle rispettive ed intrinseche proprietà fenomeniche. Inoltre, anche coloro che sono a favore di un approccio di natura fenomenologica rispetto al contenuto concettuale (Horgan e Tiensen, Loar, Pitt, Searle, Siewert) sembrano propendere per un'individuazione internalistica del contenuto (se non del riferimento) di tali stati.

7. Individualismo ed externalismo

Gli externalisti sono generalmente coloro che, come i teorici dell'informazione, ritengono che il contenuto delle rappresentazioni mentali sia determinato (almeno in parte) dal rapporto del singolo con l'ambiente (naturale o sociale) in cui esso si muove (Burge, 1979, 1986b; McGinn, 1977; Putnam, 1975), mentre gli individualisti (cfr. Putnam, 1975; Fodor, 1981b) ritengono, come alcuni sostenitori delle teorie funzionali, che il contenuto rappresentazionale sia determinato esclusivamente dalle proprietà intrinseche di un individuo.

La questione è di notevole importanza dal momento che si presume che le argomentazioni psicologiche, relative alla psi-

cologia del senso comune da una parte e quella scientifica dall'altra, siano di natura causale o basate su un contenuto. (Convinzioni e desideri provocano un determinato comportamento in quanto possiedono un determinato contenuto che spinge l'individuo ad agire in un determinato modo. Ad esempio il desiderio di una birra e la convinzione che questa si trovi nel frigorifero e che il frigorifero a sua volta sia in cucina, potrebbero spiegare perché ci si alza e si va in cucina).

Tuttavia se una rappresentazione mentale possiede un contenuto specifico, ciò è dovuto a fattori ad essa esterni; non è chiaro però come possedere tale contenuto possa determinare il suo potere causale, che certamente deve essere intrinseco (cfr. Stich, 1983; Fodor, 1982; 1987, 1994). Chi accetta le argomentazioni classiche dell'esternalismo sostiene che i fattori interni determinano un componente del contenuto della rappresentazione mentale e afferma che queste ultime possiedono un contenuto «ristretto» (cioè determinato da fattori intrinseci) e uno «ampio» (determinato dal contenuto ristretto e da fattori intrinseci). (Questa distinzione potrebbe essere applicata alle rappresentazioni sub-personali della scienza cognitiva così come a quelle relative alla psicologia del senso comune. Cfr. Von Eckardt, 1993: 189).

Al contenuto ristretto sono state date diverse interpretazioni. Ad esempio Putnam (1975), Fodor (1982: 114; 1994: 39ff), e Block (1986: 627) ritengono che sia simile al contenuto *de dicto*. Secondo questa interpretazione il contenuto ristretto è indipendente dal contesto e può essere direttamente espresso. Al contrario Fodor e Block sostengono che sia assolutamente inesprimibile. In questo modo il contenuto ristretto è visto come una sorta di proto-contenuto o determinante del contenuto che può essere specificato solo indirettamente attraverso specificazioni di associazioni contesto/contenuto ampio. Secondo le due interpretazioni, i contenuti ristretti sono definiti come funzioni contesto/contenuto ampio.

Il contenuto ristretto della rappresentazione è determinato dalle sue proprietà intrinseche e dalla propria struttura sintat-

tica o ruolo inferenziale o computazionale intramentale (o la sua fenomenologia, cfr. Searle, 1992; Siewert, 1998; Pitt).

Burge (1986b) ha sostenuto che le preoccupazioni di natura causazionale nei confronti dell'individuazione externalista del contenuto psicologico e l'introduzione di concetto ristretto siano fuorvianti. Fodor (1994, 1998) ha recentemente insistito nell'affermare che la psicologia scientifica potrebbe non aver bisogno del contenuto ristretto per supportare le argomentazioni naturalistiche (causali) dell'azione e della cognizione umana, dal momento che i casi che sono stati loro sottoposti per poter essere risolti (i casi della «terra gemella» e quelli di Frege) sono impossibili ed irrilevanti dal punto di vista nomologico in quanto eccezioni di leggi psicologiche non ristrette.

8. *La teoria computazionale della mente*

La versione contemporanea prevalente della teoria rappresentazionale della mente, ovvero, la teoria computazionale della mente (TCM), sostiene che il cervello sia una sorta di computer e che i processi mentali non siano altro che computazioni. Secondo la TCM, gli stati cognitivi sono costituiti da relazioni computazionali tra rappresentazioni di vario genere ed i processi cognitivi sono sequenze di tali stati.

La TCM sviluppa la teoria rappresentazionale della mente (TRM) tentando di spiegare tutti gli stati ed i processi psicologici in termini di rappresentazioni mentali. Per elaborare dettagliate teorie empiriche sulla cognizione umana ed animale e sviluppare modelli per i processi cognitivi applicabili a sistemi di elaborazione di informazioni artificiali, i cognitivisti hanno proposto diversi tipi di rappresentazioni mentali. Mentre alcune di queste possono essere adattate agli stati della psicologia del senso comune, per altre rappresentazioni, cosiddette «sub-personali» o «sub-doxastiche», ciò è impossibile. Sebbene molti filosofi siano concordi nell'affermare che la TCM possa spiegare nel miglior modo possibile la cognizione ed il comportamento, non tutti ritengono invece che queste

possano giustificare le argomentazioni della psicologia del senso comune della TRM pre-scientifica.

Ad esempio, secondo la «teoria sintattica della mente» di Stich, le teorie computazionali degli stati psicologici dovrebbero riguardare gli stessi solo in funzione delle proprietà formali degli oggetti con i quali questi stati sono in relazione. Tuttavia, per molti cognitivisti è fondamentale la rilevanza esplicativa del contenuto (Fodor, 1981a; Pylyshyn, 1984; Von Eckardt, 1993). I principi cardine del pensiero cognitivista sono: i processi sono computazioni, le computazioni sono sequenze disciplinate da regole di oggetti valutabili da un punto di vista semantico e tali regole vengono applicate ai simboli in virtù del loro contenuto.

Le argomentazioni nell'ambito della scienza cognitiva fanno riferimento a diversi tipi di rappresentazioni mentali, tra cui, ad esempio, i «modelli mentali» di Johnson-Laird (1983), la «configurazione retinica», lo «sketch primario» e lo «sketch a 2-D e 1/2» di Marr (1982), i «frame» di Minsky (1974), le «strutture sub-simboliche» di Smolensky (1989), la «struttura simil-figurativa» di Kosslyn (1980) e le «configurazioni simboliche» di Tye (1991), oltre a quelle rappresentazioni adatte a spiegare gli stati psicologici del senso comune. Tra gli altri fenomeni mentali, le argomentazioni computazionali comprendono credenze (Fodor, 1975; Field, 1978), percezione visiva (Marr, 1982; Osherson et al., 1990), razionalità (Newell e Simon, 1972; Fodor, 1975; Johnson-Laird e Wason, 1977), apprendimento e uso del linguaggio (Chomsky, 1965; Pinker, 1989) e comprensione musicale (Lerdahl e Jackendoff, 1983).

I sostenitori della TCM non concordano invece su ciò che concerne la realizzazione di rappresentazioni a livello personale (es. pensieri) e i processi (es. interferenze) cerebrali. A scontrarsi all'interno di questo dibattito sono i sostenitori dell'«architettura classica» e dell'«architettura connessionista».

I classicisti (Turing, 1950; Fodor, 1975; Fodor e Pylyshyn, 1988; Marr, 1982; Newell e Simon, 1976) ritengono che le rappresentazioni siano strutture simboliche che generalmente possiedono componenti valutabili da un punto di vista se-

mantico i cui processi mentali sono delle loro manipolazioni disciplinate da regole sensibili alla corrispondente struttura costitutiva. I connessionisti (McCulloch e Pitts, 1943; Rumelhart, 1989; Rumelhart e McClelland, 1986; Smolensky, 1988) sostengono che le rappresentazioni mentali vengono realizzate a partire da pattern di attivazione in una rete di processori semplici («nodi») e che i processi mentali consistono nell'attivazione diffusa di tali pattern. Gli stessi nodi non possono essere valutati da un punto di vista semantico così come i componenti dei pattern. (Sebbene esistano alcune versioni connessioniste, ad esempio quella dei «localisti», secondo cui ogni singolo nodo possiede delle proprietà semantiche (Ballard, 1986; Ballard e Hayes, 1984), tuttavia, si può sostenere che le teorie localiste non sono né definitive né rappresentative del programma connessionista. Cfr. Smolensky, 1988, 1991; Chalmers 1993).

Uno dei presupposti delle teorie classiciste è che il pensiero sembra condividere con il linguaggio alcune proprietà. L'«ipotesi del linguaggio del pensiero» (*Language of Thought Hypothesis*) di Fodor, secondo cui il sistema dei simboli mentali che costituisce la base neurale del pensiero ha la stessa struttura di un linguaggio, fornisce una versione ben elaborata dell'approccio classico applicato alla psicologia del senso comune. (Cfr. anche Marr, 1982, per l'applicazione dell'approccio classico alla psicologia scientifica). Secondo l'ipotesi del linguaggio del pensiero, un numero potenzialmente illimitato di stati mentali rappresentazionali complessi è generato da un numero limitato di stati rappresentazionali primitivi secondo regole di formazione ricorsiva.

Questa struttura combinatoria rappresenta le proprietà produttive e sistematiche del sistema delle rappresentazioni mentali. Come nel caso dei linguaggi simbolici, incluso anche quello naturale (sebbene Fodor non ritenga che l'ipotesi del linguaggio del pensiero spieghi esclusivamente le capacità linguistiche e che solo le creature dotate di parola dispongano di una tale architettura cognitiva), queste proprietà del pensiero vengono spiegate facendo riferimento al contenuto delle unità rappresentazionali e alla loro capacità di combi-

narsi dando vita a complessi ricchi di contenuto. Ovvero la semantica del linguaggio e quella del pensiero sono di natura computazionale: il contenuto di una rappresentazione complessa è determinato dai contenuti dei suoi componenti e dalla loro configurazione strutturale.

Le teorie connessioniste muovono principalmente dalla mera considerazione dell'architettura del cervello che è apparentemente costituita da reti stratificate di neuroni interconnessi. Essi sostengono che questo tipo di architettura non sia adatta ad eseguire computazioni seriali classiche e che, nel cervello, l'elaborazione è distribuita principalmente in parallelo. Inoltre gli elementi la cui manipolazione innesca la computazione nelle reti connessioniste (soprattutto quelle tra nodi) non sono composizionali e né valutabili semanticamente poiché seguono un approccio classico. Questo contrasto con il computazionalismo classico è caratterizzato dal fatto che molti sostengono che, rispetto alla computazione, la rappresentazione è distribuita in modo da opporsi a quella locale: una rappresentazione si definisce locale se è fondamentale da un punto di vista computazionale, mentre è considerata distribuita se non lo è. (Ciò può essere spiegato anche in altri termini dicendo che per i classicisti le rappresentazioni sono computazionalmente atomiche, mentre per i connessionisti non lo sono.)

I connessionisti ritengono comunque che l'elaborazione dell'informazione come avviene nelle reti connessioniste ricordi molto da vicino alcune caratteristiche del reale funzionamento cognitivo umano. Ad esempio, mentre da un punto di vista classico l'apprendimento implica la formulazione e la dimostrazione dell'ipotesi, secondo il modello connessionista esso consiste nello sviluppo della distribuzione del «peso» (forza) sulle connessioni tra nodi e generalmente non implica la formulazione dell'ipotesi sulle condizioni di identità degli oggetti della conoscenza. La rete connessionista viene «addestrata» mediante un'esposizione ripetuta agli oggetti che deve imparare a distinguere; nonostante tali debbano essere esposte agli oggetti più frequentemente rispetto agli esseri umani, ciò sembra riprodurre in maniera migliore almeno una delle

proprietà dell'apprendimento umano (cfr. l'esempio del sonar in Churchland, 1989).

Inoltre, in conseguenza ad una lesione la riduzione della funzionalità di tali reti è graduale, non improvvisa come nel caso del processore di informazione classico, e quindi riproduce con più precisione la perdita della funzione cognitiva dell'essere umano che si verifica in seguito ad una lesione cerebrale. Si ritiene anche che in alcuni casi il sistema connessionista mostri un certo tipo di flessibilità in risposta a situazioni nuove tipiche della cognizione umana, situazioni in cui i sistemi classici si rivelerebbero «fragili» e «delicati».

Alcuni filosofi sostengono che il connessionismo non implichi nessun tipo di atteggiamento proposizionale. Ramsey, Stich e Gordon (1990) hanno affermato che se i modelli connessionisti sono fondamentalmente corretti, ne deriva che non esistono stati rappresentazionali discreti così come la psicologia del senso comune e la scienza cognitiva li concepisce. Tuttavia altri (Smolensky, 1989) ritengono che, nelle reti neurali, alcuni tipi di pattern ad alto livello di attività possano approssimativamente essere identificati con gli stati rappresentazionali della psicologia del senso comune. Altri ancora (Fodor e Pylyshyn, 1988; Heil, 1991; Horgan e Tienson, 1996) affermano che la rappresentazione dello stile del linguaggio del pensiero è generalmente necessaria e realizzabile all'interno di architetture connessioniste. (MacDonald & MacDonald, 1995, raccoglie importanti documenti aggiornati sul dibattito classicista/connessionista e fornisce anche materiale introduttivo, vedi anche Von Eckardt, 2004).

Mentre Stich (1983) accetta la tesi secondo cui i processi mentali sono di natura computazionale, rifiuta quella che definisce le computazioni come sequenze di rappresentazioni mentali, altri invece accettano il concetto di rappresentazione mentale, mentre rifiutano la tesi secondo cui la «teoria computazionale della mente» fornisce una corretta descrizione degli stati e dei processi mentali.

Van Gelder (1995) rifiuta la tesi secondo cui i processi psicologici sono di natura computazionale. Afferma che i sistemi cognitivi sono dinamici e che gli stati cognitivi non so-

no in relazione ai simboli mentali, ma sono stati quantificabili di sistemi complessi composti (nel caso degli esseri umani) da un sistema nervoso, un corpo ed un ambiente in cui questi sono inseriti. I processi cognitivi non sono sequenze disciplinate da regole di stati simbolici discreti, ma di stati totali in evoluzione di sistemi dinamici determinati da stati continui, simultanei e reciprocamente determinanti dei componenti dei sistemi. La rappresentazione in un sistema dinamico è essenzialmente un'informazione teoretica, sebbene i «trasmettitori» delle informazioni (*bearers of information*) non siano simboli, ma variabili o parametri dello stato. (Vedi anche Port e Van Gelder, 1995; Clark, 1997a, 1997b).

Horst (1996), dal canto suo, sostiene che sebbene i modelli computazionali possano essere utili nell'ambito della psicologia scientifica, non aiutano ad interpretare da un punto di vista filosofico l'intenzionalità degli stati mentali del senso comune. La teoria computazionale della mente tenta di ridurre l'intenzionalità di tali stati all'intenzionalità dei simboli mentali con cui essi sono in relazione. Horst afferma però che il concetto di contenuto simbolico è strettamente legato a quello di convenzione ed intenzione. Quindi la TCM stessa si ritrova in un circolo vizioso: le proprietà fondamentali che dovrebbero essere ridotte sono (tacitamente) attratte dalla riduzione.

9. Pensiero e Linguaggio

Dire che un oggetto mentale possiede delle proprietà semantiche significa dire che questo si riferisce alla verità o alla falsità di uno o più oggetti o, più semplicemente, che questo sia vero o falso. Immaginiamo che la mia supposizione sia: i leopardi respirano. Sto pensando ai leopardi, e se ciò che sto pensando di essi (che respirano) è vero, allora il mio pensiero è vero. Secondo la TRM questi stati sono definiti come relazioni tra agenti e rappresentazioni mentali. Pensare che i leopardi respirino significa, in un certo senso, elaborare una rappresentazione mentale, il cui contenuto afferma che i leopardi respirano. Secondo questa teoria, le proprietà semantiche

degli stati mentali non sono altro che proprietà semantiche delle rappresentazioni a cui essi sono relazionati.

Sembra che gli atti linguistici condividano tali proprietà con gli stati mentali. Supponiamo che io pensi che i leopardi respirino, e se ciò che dico di loro (che respirano) è vero, allora la mia affermazione è vera. Ora dire che i leopardi respirano significa (parzialmente) formulare una frase il cui significato è che i leopardi respirano. Secondo molti filosofi, le proprietà semantiche di una espressione linguistica derivano dagli stati mentali intenzionali che convenzionalmente esprimono. Secondo questa teoria, le proprietà semantiche delle espressioni linguistiche sono proprietà semantiche di *relata* mentali di stati che convenzionalmente esprimono.

Altri tuttavia – per esempio Davidson (1975, 1982) – hanno sostenuto che il tipo di pensiero di cui sono dotati gli esseri umani non sarebbe possibile senza il linguaggio, sicché la dipendenza dovrebbe essere rovesciata, o in qualche modo reciproca. (Vedi anche Sellars, 1956; ma vedi anche Martin, 1987, in difesa dell'asserzione secondo cui il pensiero è possibile anche senza il linguaggio. Vedi anche Chisholm e Sellars, 1958). Schiffer (1987) ha successivamente riflettuto sul successo di quella che lui stesso ha definito «semantica basata sull'intenzione» (*Intention Based Semantics*).

È stato anche sostenuto che oltre ad avere come riferimento proprietà come referenza, condizione di verità e verità – cosiddette proprietà estensionali – le espressioni dei linguaggi naturali possiedono anche proprietà *intenzionali*: ad esempio due espressioni possono avere identici riferimenti, condizione di verità o valori di verità, ma esprimere proprietà e proposizioni differenti (Frege, 1892/1997). Se le proprietà semantiche delle espressioni del linguaggio naturale derivano dai pensieri e dal concetto che esprimono, allora sarebbe opportuna una distinzione analoga per quanto riguarda le rappresentazioni mentali.

Bibliografia

- ALMOG, J., PERRY, J., WETTSTEIN, H. (a cura di), *Themes from Kaplan*, New York, Oxford University Press, 1989.
- ARISTOTELE, *De Anima*, in *The Complete Works of Aristotle: The Revised Oxford Translation*, Oxford, Oxford University Press, 1984.
- BALLARD, D.H., *Cortical Connections and Parallel Processing: Structure and Function*, in «The Behavioral and Brain Sciences», 1986, pp. 67-120.
- BALLARD, D.H. and Hayes, P.J., *Parallel Logical Inference*, in «Proceedings of the Sixth Annual Conference of the Cognitive Science Society», Rochester, New York, 1984.
- BEANEY, M., (a cura di), *The Frege Reader*, Oxford, Blackwell Publishers, 1997
- BERKELEY, G., *Principles of Human Knowledge*, in M.R. Ayers (a cura di), *Berkeley: Philosophical Writings*, London, Dent, 1975.
- BLOCK, N. (a cura di), *Readings in Philosophy of Psychology, Vol. 2*, Cambridge, Mass., Harvard University Press, 1981.
- (a cura di), *Imagery*, Cambridge, Mass., The MIT Press, 1982.
- , *Mental Pictures and Cognitive Science*, in «Philosophical Review» 93, 1983, pp. 499-542.
- , *Advertisement for a Semantics for Psychology*, in P.A. FRENCH, T.E. UEHLING, H.K. WETTSTEIN (a cura di), «MidwestStudies in Philosophy», Vol. X, Minneapolis, University of Minnesota Press, 1986, pp. 615-678.
- , *Mental Paint and Mental Latex*, in E. VILLANUEVA (a cura di), *Philosophical Issues, 7: Perception*, 1996, 19-49.
- BOGHOSSIAN, P.A., *Content*, in J. KIM, E. SOSA, (a cura di), *A Companion to Metaphysics*, Oxford, Blackwell Publishers Ltd., 1995, pp. 94-96.
- BURGE, T., *Individualism and the Mental*, in P.A. French, T.E. Uehling and H.K. Wettstein (a cura di), «Midwest Studies in Philosophy» Vol. IV, Minneapolis University of Minnesota Press, 1986, pp. 73-121.

- CHALMERS, D. *The Conscious Mind*, New York, Oxford University Press, 1996.
- , *Connectionism and Compositionality: Why Fodor and Pylyshyn Were Wrong*, in «Philosophical Psychology», 6, 1993, pp. 305-319.
- CHISHOLM, R., SELLARS, W. *The Chisholm-Sellars Correspondence on Intentionality*, in H. FEIGL, M. SCRIVEN, G. MAXWELL (a cura di), «Minnesota Studies in the Philosophy of Science», Vol. II, Minneapolis, University of Minnesota Press, 1958, pp. 529-539.
- CHOMSKY, N. *Aspects of the Theory of Syntax*, Cambridge, Mass., The MIT Press, 1965.
- CHURCHLAND, P.M., *Eliminative Materialism and the Propositional Attitudes*, in «Journal of Philosophy», 78, 1981, pp. 67-90.
- , *On the Nature of Theories: A Neurocomputational Perspective*, in W. SAVAGE (a cura di), «Scientific Theories: Minnesota Studies in the Philosophy of Science», Vol. 14, Minneapolis, University of Minnesota Press, 1989, pp. 59-101.
- DAVIDSON, D., *Radical Interpretation*, in «Dialectica», 27, 1973, pp. 313-328.
- , *Belief and the Basis of Meaning*, in «Synthese», 27, 1974, pp. 309-323.
- , *Thought and Talk*, in S. GUTTENPLAN (a cura di), *Mind and Language*, Oxford, Clarendon Press, 1975, pp. 7-23.
- , *Rational Animals*, in «Dialectica», 4, 1982, pp. 317-327.
- DENNETT, D. *Content and Consciousness*, London, Routledge and Kegan Paul, 1969.
- , *The Nature of Images and the Introspective Trap*, in BLOCK 1981, pp. 128-134.
- , *The Intentional Stance*, Cambridge, Mass., The MIT Press, 1987.
- , *True Believers: The Intentional Strategy and Why it Works*, in Dennett 1987 (1987a), pp. 13-35.
- , *Reflections: Real Patterns, Deeper Facts, and Empty Questions*, in DENNETT 1987 (1987b), pp. 37-42.

La rappresentazione mentale

- , *Quining Qualia*, in A.J. MARCEL, E. BISIACH (a cura di), *Consciousness in Contemporary Science*, Oxford, Clarendon Press, 1988, pp. 42-77.
- , *Real Patterns*, in «The Journal of Philosophy», LXXXVII, 1991, pp. 27-51.
- DEVITT, M., *Coming to Our Senses: A Naturalistic Program for Semantic Localism*, Cambridge, Cambridge University Press, 1996.
- DRETSKE, F. *Seeing and Knowing*, Chicago, The University of Chicago Press, 1969.
- , *Knowledge and the Flow of Information*, Cambridge, Mass., The MIT Press 1981.
- , *Explaining Behavior: Reasons in a World of Causes*, Cambridge, Mass., The MIT Press, 1988.
- , *Naturalizing the Mind*, Cambridge, Mass., The MIT Press, 1994.
- EVANS, G. *The Varieties of Reference*, Oxford, Oxford University Press, 1982.
- FIELD, H., *Mental representation*, in «Erkenntnis», 13, 1978, pp. 9-61.
- FLANAGAN, O., *Consciousness Reconsidered*, Cambridge, Mass., The MIT Press, 1992.
- FODOR, J.A., *The Language of Thought*, Cambridge, Mass., Harvard University Press, 1975.
- , *Propositional Attitudes*, in «The Monist», 61, 1978, pp. 501-523.
- , *Representations*, Cambridge, Mass., The MIT Press, 1981.
- , *Introduction*, in FODOR 1981 (1981a), pp. 1-31.
- , *Methodological Solipsism Considered as a Research Strategy in «Cognitive Psychology»* in FODOR 1981 (1981b), pp. 225-253.
- , *The Present Status of the Innateness Controversy*, in FODOR 1981 (1981c), pp. 257-316.
- , *Cognitive Science and the Twin-Earth Problem*, in «Notre Dame Journal of Formal Logic», 23, 1982, pp. 98-118.
- , *Psychosemantics*, Cambridge, Mass., The MIT Press, 1987.
- , *A Theory of Content and Other Essays*, Cambridge, Mass., The MIT Press, 1990a.

- , *Psychosemantics or: Where Do Truth Conditions Come From?* in W.G. LYCAN (a cura di), *Mind and Cognition: A Reader*, Oxford, Blackwell Publishers, 1990, 1990b, pp. 312-337.
- , *The Elm and the Expert*, Cambridge, Mass., The MIT Press, 1994.
- FODOR, J.A., PYLYSHYN, Z., *How Direct is Visual Perception?: Some Reflections on Gibson's 'Ecological Approach'*, in «Cognition», 9, 1981, pp. 207-246.
- , *Connectionism and Cognitive Architecture: A Critical Analysis*, in «Cognition», 28, 1988, pp. 3-71.
- FREGE, G. *The Foundations of Arithmetic*, trad. di J.L. Austin, New York, Philosophical Library, 1953.
- , *On Sinn and Bedeutung* (1892), in Beany 1997, pp. 151-171.
- , *Thought* (1918), in BEANY 1997, pp. 325-345.
- GEACH, P. *Mental Acts: Their Content and Their Objects*, London, Routledge & Kegan Paul, 1957.
- GOLDMAN, A., *The Psychology of Folk Psychology*, in «Behavioral and Brain Sciences», 16, 1993, pp. 15-28.
- GOODMAN, N., *Languages of Art* (2° ed.), Indianapolis, Hackett, 1976.
- GRICE, H.P., *Meaning*, in «Philosophical Review», 66, 1957, pp. 377-388; ristampato in *Studies in the Way of Words*, Cambridge, Mass., Harvard University Press, 1989, pp. 213-223.
- HARMAN, G. *Thought*, Princeton, Princeton University Press, 1973.
- , *(Non-Solipsistic) Conceptual Role Semantics*, in E. LEPORE (a cura di), *New Directions in Semantics*, London, Academic Press, 1987, pp. 55-81.
- , *The Intrinsic Quality of Experience* in J. Tomberlin (a cura di), *Philosophical Perspectives 4: Action Theory and Philosophy of Mind*, Atascadero, Ridgeview Publishing Company, 1990, pp. 31-52.
- HAUGELAND, J., *Analog and analog*, in «Philosophical Topics», 12, 1981, pp. 213-226.

- HEIL, J., *Being Indiscrete*, in J. Greenwood (a cura di), *The Future of Folk Psychology*, Cambridge, Cambridge University Press 1991, pp. 120-134.
- HORGAN, T., TIENSON, J. *Connectionism and the Philosophy of Psychology*, Cambridge, Mass., The MIT Press 1996.
- HORST, S., *Symbols, Computation, and Intentionality*, Berkeley, University of California Press 1996.
- HUME, D., *A Treatise of Human Nature*, a cura di L.A. Selby-Bigge, revisionato da P.H. Nidditch, Oxford, Oxford University Press 1978.
- JACKENDOFF, R. *Computation and Cognition*, Cambridge, Mass., The MIT Press 1987.
- JOHNSON-LAIRD, P.N., *Mental Models*, Cambridge, Mass., Harvard University Press 1983.
- JOHNSON-LAIRD, P.N., WASON, P.C. *Thinking: Readings in Cognitive Science*, Cambridge University Press 1977.
- KAPLAN, D., *Demonstratives*, in ALMOG, PERRY, WETTSTEIN 1989, pp. 481-614.
- KOSSLYN, S.M. *Image and Mind*, Cambridge, Mass., Harvard University Press 1980.
- , *The Medium and the Message in Mental Imagery*, in Block 1982, pp. 207-246.
- , *Ghosts in the Mind's Machine*, New York, W.W. Norton & Co. 1983.
- KOSSLYN, S.M., POMERANTZ, J.R., *Imagery, Propositions, and the Form of Internal Representations*, in «Cognitive Psychology», 9, 1977, pp. 52-76.
- LEEDS, S., *Qualia, Awareness, Sellars*, in «Noûs», XXVII, 1993, pp. 303-329.
- LERDAHL, F. and JACKENDOFF, R., *A Generative Theory of Tonal Music*, Cambridge, Mass., The MIT Press 1983.
- LEVINE, J., *On Leaving Out What It's Like*, in M. DAVIES, G. HUMPHREYS (a cura di), *Consciousness*, Oxford, Blackwell Publishers, 1993, pp. 121-136.
- LEVINE, J., *On What It Is Like to Grasp a Concept*, in E. VILLANUEVA (a cura di), *Philosophical Issues 6: Content*, Atascadero, Ridgeview Publishing Company, 1995, pp. 38-43.

- LEWIS, D., *Analog and Digital*, in «Noûs», 5, 1971, pp. 321-328.
- , *Radical Interpretation*, in «Synthese», 23, 1974, pp. 331-344.
- , *Philosophical Papers, Vol. I*, New York, Oxford University Press 1983.
- LOAR, B., *Mind and Meaning*, Cambridge, Cambridge University Press, 1981.
- LOCKE, J., *An Essay Concerning Human Understanding*, a cura di P.H. Nidditch, Oxford, Oxford University Press, 1978.
- MACDONALD, C., MacDONALD, G., *Connectionism: Debates on Psychological Explanation*, Oxford, Blackwell Publishers, 1995.
- MARR, D., *Vision*, New York, W.H. Freeman and Company, 1982.
- MARTIN, C.B., *Proto-Language*, in «Australasian Journal of Philosophy», 65, 1987, pp. 277-289.
- MCCULLOCH, W.S., PITTS, W., *A Logical Calculus of the Ideas Immanent in Nervous Activity*, in «Bulletin of Mathematical Biophysics», 5, 1943, pp. 115-33.
- MCGINN, C., *The Structure of Content*, in A. Woodfield (a cura di), *Thought and Content*, Oxford, Oxford University Press, 1982, pp. 207-258.
- , *Mental Content*, Oxford, Blackwell Publishers, 1989.
- , *Content and Consciousness*, in C. McGinn, *The Problem of Consciousness*, Oxford, Blackwell Publishers 1992, pp. 23-43.
- MILLIKAN, R. *Language, Thought and other Biological Categories*, Cambridge, Mass., The MIT Press, 1984.
- NEWELL, A., SIMON, H.A., *Human Problem Solving*, New York, Prentice-Hall, 1972.
- , *Computer Science as Empirical Inquiry: Symbols and Search*, in «Communications of the Association for Computing Machinery», 19, 1976, pp. 113-126.
- OSHERSON, D.N., KOSSLYN, S.M., HOLLERBACH, J.M., *Visual Cognition and Action: An Invitation to Cognitive Science, Vol. 2*, Cambridge, Mass., The MIT Press, 1990.
- PAPINEAU, D., *Reality and Representation*, Oxford, Blackwell Publishers, 1987.

- PEACOCKE, C., *Sense and Content*, Oxford, Clarendon Press 1983.
- , *Perceptual Content*, in ALMOG, PERRY e WETTSTEIN 1989 (1989), pp. 297-329.
- , *Scenarios, Concepts and Perception*, in T. CRANE (a cura di), *The Contents of Experience*, Cambridge, Cambridge University Press 1992, pp. 105-35.
- PINKER, S., *Learnability and Cognition*, Cambridge, Mass., The MIT Press 1989.
- PORT, R., VAN GELDER, T., *Mind as Motion: Explorations in the Dynamics of Cognition*, Cambridge, Mass., The MIT Press 1995.
- PUTNAM, H., *The Meaning of 'Meaning'*, in «Philosophical Papers», Vol. 2, Cambridge, Cambridge University Press, 1975, pp. 215-71.
- PYLYSHYN, Z., *The Rate of 'Mental Rotation' of Images: A Test of a Holistic Analogue Hypothesis*, in «Memory and Cognition», 7, 1979, pp. 19-28.
- , *Imagery and Artificial Intelligence*, in BLOCK 1981 (1981a), pp. 170-194.
- , *The Imagery Debate: Analog Media versus Tacit Knowledge*, in «Psychological Review», 88 (1981b), pp. 16-45.
- , *Computation and Cognition*, Cambridge, Mass., The MIT Press 1984.
- RAFFMAN, D., *The Persistence of Phenomenology*, in T. METZINGER (a cura di), *Conscious Experience*, Paderborn, Schöningh/Imprint Academic 1995, pp. 293-308.
- RAMSEY, W., STICH, S., GARON, J., *Connectionism, Eliminativism and the Future of Folk Psychology*, in «Philosophical Perspectives», 4, 1990, pp. 499-533.
- REY, G., *Introduction: What Are Mental Images?* in BLOCK 1981, pp. 117-127.
- , *Sensations in a Language of Thought*, in E. VILLANEUVA (a cura di), *Philosophical Issues 1: Consciousness*, Atascadero: Ridgeview Publishing Company 1991, pp. 73-112.
- RUMELHART, D.E., *The Architecture of the Mind: A Connectionist Approach*, in M.I. POSNER (a cura di),

- Foundations of Cognitive Science*, Cambridge, Mass., The MIT Press 1989, pp. 133-159.
- RUMELHART, D.E., MCCELLAND, J.L., *Parallel Distributed Processing, Vol. I*, Cambridge, Mass., The MIT Press, 1986.
- SCHIFFER, S., *Remnants of Meaning*, Cambridge, Mass., The MIT Press, 1987.
- , *Introduction to the Paperback Edition*, in *Meaning*, Oxford, Clarendon Press 1988, pp. XI-XXIX.
- SEARLE, J.R. *Intentionality*, Cambridge, Cambridge University Press, 1983.
- , *The Rediscovery of the Mind*, Cambridge, Mass., The MIT Press, 1990.
- SELLARS, W., *Empiricism and the Philosophy of Mind*, in K. Gunderson (a cura di), «Minnesota Studies in the Philosophy of Science», Vol. I, Minneapolis: University of Minnesota Press, 1956, pp. 253-329.
- SHEPARD, R.N., COOPER, L., *Mental Images and their Transformations*, Cambridge, Mass., The MIT Press, 1982.
- SHOEMAKER, S., *Qualities and Qualia: What's in the Mind?*, in «Philosophy and Phenomenological Research», 50, 1990, pp. 109-31.
- SMOLENSKY, P., *On the Proper Treatment of Connectionism*, in «Behavioral and Brain Sciences», 11, 1988, pp. 1-74.
- , *Connectionist Modeling: Neural Computation/Mental Connections*, in L. NADEL, L.A. COOPER, P. CULICOVER, R.M. HARNISH (a cura di), *Neural Connections, Mental Computation*, The MIT Press, 1989, pp. 49-67.
- , *Connectionism and the Language of Thought*, in B. LOEWER, G. REY (a cura di), *Meaning in Mind: Fodor and His Critics*, Oxford, Basil Blackwell Ltd, 1991, pp. 201-227.
- STERELNY, K., *Fodor's Nativism*, in «Philosophical Studies», 55, 1989, pp. 119-141.
- STICH, S. 1978., *Autonomous Psychology and the Belief-Desire Thesis*, in «The Monist», 61, 1978, pp. 573-591.
- , *From Folk Psychology to Cognitive Science*, Cambridge, Mass., The MIT Press, 1983.
- STRAWSON, G. *Mental Reality*, Cambridge, Mass., The MIT Press, 1994.

- TYE, M. *The Imagery Debate*, Cambridge, Mass., The MIT Press, 1991.
- , *Ten Problems of Consciousness*, Cambridge, Mass., The MIT Press, 1995.
- , *Consciousness, Color, and Content*, Cambridge, Mass., The MIT Press, 2000.
- VAN GELDER, T., *What Might Cognition Be, if not Computation?*, in «Journal of Philosophy», XCI, 1995, pp. 345-381.
- VON ECKARDT, B., *What Is Cognitive Science?*, Cambridge, Mass., The MIT Press, 1993.
- , *Connectionism and the Propositional Attitudes*, in C. EMELING, D.M. JOHNSON (a cura di), *The Mind as a Scientific Object: Between Brain and Culture*, 1993
- WITTGENSTEIN, L., *Philosophical Investigations*, trad. di G.E.M. Anscombe, Oxford, Blackwell Publishers, 1953.

LA TEORIA COMPUTAZIONALE DELLA MENTE⁵¹

Steven Horst

Negli ultimi trent'anni è accaduto spesso di sentir paragonare la mente ad un computer digitale. Questo saggio tratta un particolar punto di vista filosofico secondo il quale la mente corrisponde ad un computer (nel senso specifico di «macchina che computa»), ed il pensiero ad una sorta di computazione. Questa teoria, che verrà chiamata «teoria computazionale della mente» (CTM, Computational Theory of Mind), deve perciò essere distinta dagli altri numerosi tentativi di mettere in relazione mente e computazione, inclusi: (a) i tentativi di modellare le caratteristiche della mente utilizzando tecniche di modellamento computazionale, e (b) l'utilizzo di una o più caratteristiche del modello di produzione dei computer (come il concetto di programma memorizzato, o la differenza tra hardware e software) come una metafora per la comprensione di alcune caratteristiche della mente. Ciò è strettamente correlato con la CTM elaborata da Hilary Putnam (1961) e più accuratamente perfezionata da Jerry Fodor (1975, 1980, 1987, 1993). Il significato di «computer» e «computazione» qui impiegato è tecnico. Le funzioni principali di questa nuova teoria dovranno essere in grado di spiegare: (a) il significato tecnico di «computazione» in questione; (b) le modalità in cui debba essere applicato alla mente; (c) i problemi filosofici che questa visione della mente do-

⁵¹ S. Horst, *The Computational Theory of Mind*, in «Stanford Encyclopedia of Philosophy», <http://plato.stanford.edu/entries/computational-mind/>

vrebbe risolvere; (d) le maggiori critiche rivolte a questa teoria.

1. *Tesi principali*

La teoria computazionale della mente combina una serie di ragionamenti con una serie di stati mentali. Più recentemente essa è stata denominata anche «teoria rappresentazionale della mente» (TRM). Secondo questa tesi, stati intenzionali come credenze e desideri sono relazioni tra un pensante e delle rappresentazioni simboliche del contenuto degli stati: ad esempio, credere che ci sia un gatto sul tappeto significa trovarsi in una particolare relazione funzionale (caratteristica dell'attitudine a credere) con una rappresentazione mentale simbolica il cui valore semantico è «c'è un gatto sul tappeto»; sperare che ci sia un gatto sul tappeto significa invece trovarsi in una relazione funzionale differente (caratteristica dell'attitudine a *sperare* piuttosto che a credere) rispetto ad una rappresentazione mentale simbolica che ha lo stesso valore semantico.

La tesi sul ragionamento, che chiameremo «descrizione computazionale del ragionamento» (*Computational Account of Reasoning*, CAR), dipende essenzialmente dalla precedente affermazione che gli stati intenzionali implicano rappresentazioni simboliche. Secondo il CAR queste rappresentazioni hanno proprietà sia semantiche che sintattiche ed i processi del ragionamento avvengono rispondendo soltanto alla sintassi dei singoli – un tipo di processo che incontra una definizione tecnica di «computazione» conosciuta come manipolazione formale del simbolo (cioè manipolazione dei simboli secondo tecniche puramente formali e non semantiche).

1.1 *Le nozioni di «tecnica formale» e «computazione»*

Le nozioni di «manipolazione formale del simbolo» e «computazione» sono tecniche, e in definitiva derivano da discussioni matematiche sorte a cavallo tra il XIX e XX secolo. Il progetto di formalizzazione iniziò in risposta ad una crisi

che si era sviluppata in campo matematico a seguito della scoperta della possibile esistenza di geometrie plausibili che però negavano il postulato delle parallele di Euclide (secondo il quale data una linea retta R giacente su un piano e un punto P dello stesso piano non appartenente alla retta, esiste una ed una sola retta passante per P e parallela a R). L'evidente plausibilità del postulato delle parallele si basava non su qualcosa di esplicito nel sistema euclideo, ma semplicemente su intuizioni geometrico-spaziali profondamente radicate. A lungo si è supposto che tali affermazioni, intuitivamente corrette, della geometria euclidea fossero necessariamente vere, nel senso che non potevano essere logicamente negate. La scoperta, nei primi anni del XIX secolo, di possibili geometrie non concordanti con le nostre intuizioni spaziali spinse matematici come Gauss, Peano, Frege e Hilbert a cercare nuove regole al fine di disciplinare i ragionamenti matematici e ottenere così delle conclusioni fondate su assiomi specifici e su regole da essi deducibili; le intuizioni semantiche dei matematici furono, quindi, o escluse o esplicitamente codificate.

Il metodo più influente per la formalizzazione fu quello di Hilbert, che trattò ragionamenti formalizzati come un «gioco simbolico», nei quali le regole di derivazione erano espresse in termini di proprietà sintattiche (o, forse meglio, non-semantiche) dei simboli impiegati. Uno dei maggiori risultati del programma formalista fu la scoperta che grandi settori della matematica possono essere realmente formalizzati in questo modo – cioè che relazioni semantiche, ritenute fondamentali in un campo importante come la geometria, possono in effetti essere protette dalle interferenze, e rese sensibili solamente alla forma sintattica delle espressioni.

Hilbert stesso sviluppò tale progetto nell'ambito della geometria, e successivamente Whitehead e Russell estesero tale metodo all'aritmetica. Il progetto servì da modello, con minor successo, in altri campi – come ad esempio il comportamentismo in psicologia – piuttosto riduttivi rispetto alla matematica. Anche in matematica, comunque, vi sono dei limiti rispetto a quel che può essere formalizzato: i più importanti dei

quali sono stati mostrati dal teorema dell'incompletezza di Gödel.

Una seconda importante questione matematica del XIX e XX secolo fu quella di delimitare la classe di funzioni «computabili» nel senso tecnico di essere decidibili o valutabili attraverso l'applicazione di un algoritmo o di una procedura meccanica. (Esempi familiari di procedure algoritmiche sono l'addizione in colonna e le equazioni differenziali). Non tutte le funzioni matematiche si possono computare in questo modo, ma, mentre questo era noto ai matematici già nel XIX secolo, non fu prima del 1936 che Alan Turing propose di caratterizzare la classificazione delle funzioni computabili. Fu in questo contesto che egli propose la nozione di «macchina commutatrice» – cioè di una macchina che agisce analogamente ad un matematico che «computa» una funzione, ovvero che la valuta applicando una mera procedura meccanica. Turing propose che la classe delle funzioni computabili fosse equivalente alla classe di funzioni calcolabili in un numero finito di passaggi con una macchina che lui stesso aveva progettato. L'idea di fondo era che ciascuna delle operazioni sensibili solo alla sintassi potesse essere riprodotta (o forse simulata) meccanicamente. Come un matematico che segue un algoritmo formale riconosce i modelli sintattici nella sintassi, così era possibile costruire una macchina che facesse lo stesso avvalendosi di meri strumenti meccanici.

Formalizzazione e computazione sono, quindi, strettamente correlate e insieme portano alla conclusione che i ragionamenti formalizzabili possono anche essere riprodotti o simulati con il giusto tipo di macchina. Turing stesso sembra essere stato dell'opinione che una macchina che operasse in tale modo farebbe letteralmente le stesse cose che fa la computazione umana – ovvero «riprodurre» ciò che fa la mente. Ma altri autori hanno suggerito che quel che il computer-macchina fa è semplicemente una «simulazione» di quello che fa la mente umana, cioè una riproduzione delle prestazioni umane, probabilmente attraverso una serie di passaggi di tipo isomorfo a quelli dell'essere umano, che tuttavia non è

in grado di operare allo stesso modo in tutti i principali aspetti.

1.2 *Applicazione alla mente*

Come già accennato, la TCM combina una teoria rappresentazionale della mente (TRM) con il calcolo computazionale del ragionamento (CAR). La TRM è in questo caso spiegata dalla nozione di rappresentazione simbolica impiegata nella nozione tecnica di computazione: gli stati mentali sono chiamati ad essere «rappresentazionali», nel senso che includono, come costituenti, rappresentazioni simboliche con caratteristiche semantiche e sintattiche, proprio come fanno i simboli che vengono impiegati nella computazione matematica. Questa teoria, se differisce dalla prima versione moderna di rappresentazionalismo, che paragonava le idee ad immagini piuttosto che a simboli, diventa filosoficamente importante se associata al CAR. Secondo questo calcolo, il ragionamento è un processo nel quale i determinanti causali sono, nel «linguaggio del pensiero» (*language of thought*, LOT) o «mentalese», ~~caratteristiche sintattiche della simbolizzazione~~ e computazione svolgono qui in modo discutibile un'importante funzione filosofica: la formalizzazione mostra come le caratteristiche semantiche dei simboli possano (talvolta) essere codificate (*encoded*) in regole di derivazione che si basano sulla sintassi, ammettendo la possibilità di interferenze che rispettino il valore semantico che deve essere trasmesso in un modo sensibile solo alla sintassi, eludendo il bisogno, per chi fa il ragionamento, di impiegare intuizioni semantiche. In breve, la formalizzazione ci mostra il legame tra sintassi e semantica. L'idea di Turing di una macchina computatrice ci mostra come collegare la sintassi alla causazione, per cui è possibile progettare un meccanismo capace di calcolare qualsiasi funzione formalizzabile.

1.3 Scopo della teoria

Il campo più ovvio per la teoria della TCM è quello degli stati di attitudine proposizionale temporanea (*occurrent propositional attitude states*), cioè stati che accadono in momenti specifici della vita mentale di una persona e che implicano una specie di contenuto che potrebbe essere espresso con una frase proposizionale, come l'asserzione che il gatto è alla porta o il desiderio che il gatto smetta di grattare alla porta. Forse qui abbiamo i casi più plausibili di stati mentali che possono essere basati su qualcosa come le rappresentazioni mentali simboliche.

All'interno di questa classe di stati temporanei, comunque, possiamo distinguere tra stati che si verificano in giudizi espliciti e consci e stati che hanno invece luogo in maniera inconscia perché si verificano ad un «livello di processo» troppo basso perché affiori al livello della consapevolezza – per esempio i processi di percezione del contorno di un oggetto in una visione improvvisa. (Questi tipi di processi possono essere denominati «infracosci» per distinguerli da quelli consci e inconsci di Freud e Jung. Cfr Horst, 1995). Molti sostenitori della TCM applicano questa teoria non solo a livello di giudizi espliciti o desideri temporanei, ma anche ad una vasta gamma di stati infracosci. I sostenitori della TCM ne parlano spesso in termini più generici come di credenze e desideri che vengono poi portati a termine in termini disposizionali piuttosto che temporanei. Questi stati sono discutibilmente più problematici degli stati temporanei, poiché ci sono molte cose che possono essere pensate per «credere» o «desiderare» nel senso disposizionale di questi termini, ma che non potrebbero essere plausibilmente supposti ed esplicitamente rappresentati nella forma di un dato simbolo. Un'ultima questione riguardante lo scopo della teoria era quella di quanto si volessero considerare comprensibili un certo numero di stati e processi mentali. I sostenitori della TCM, come anche i critici, hanno spesso dato per scontato che la teoria fosse una sorta di descrizione di come operano i ragionamenti. Ma la questione si complica anche alla luce

della distinzione di Fodor a proposito dei processi mentali «modulari» e «globali» (Fodor, 2000). Questo punto di vista ha colto di sorpresa alcuni lettori, come anche alcuni critici; se è vero che fin dal 1970 Fodor è convinto dell'assoluta rispondenza tra la teoria e la realtà, lo stesso Fodor ammette che non gli è mai capitato di «incontrare qualcuno che credesse altrettanto; anzi, molti credono che questa teoria sia ancora lontanissima dallo spiegare come funzioni la mente» (2000: 1). Potremmo quindi passare dalle domande circa la veridicità della TCM alle domande relative alla sua veridicità come teoria per particolari tipi di processi mentali.

2. *Rilievi filosofici della teoria*

La TCM si affermò negli anni '80 come una delle più importanti teorie della mente. Questo in parte si deve all'intuizione della metafora del computer, che faceva leva sulla nozione di una tecnologia che stava rapidamente guadagnando riconoscimenti pubblici ed importanti applicazioni tecnologiche. Nel frattempo, peraltro, il computer stava condizionando la visione del funzionamento della mente, influenzando anche alcuni progetti nelle scienze cognitive (come il modello della visione di David Marr [Marr, 1983]) e nell'intelligenza artificiale, un campo in cui i ricercatori cercavano di ottenere macchine dotate – a livello di ragionamento, linguaggio, capacità di *problem solving* e percezione – di competenze di livello umano, non sempre però replicando i meccanismi con cui queste sono svolte negli esseri umani. Inoltre i sostenitori della TCM affermano anche che il computer ha effettivamente risolto alcuni problemi di natura filosofica: la sua efficacia come modello in questi campi ha contribuito in maniera sostanziale alla sua popolarità.

2.1 *Ragione e causa*

Il beneficio più importante per la TCM fu che riuscì a mostrare come il ragionamento potesse arrivare ad essere un processo causale non misterioso e come potesse anche essere

sensibile a relazioni semantiche tra i ragionamenti. Il problema di fondo era il tradizionale punto di vista secondo cui le ragioni non sono le cause. Da un lato è difficile vedere come un processo puramente formale possa avere luogo solo sulla base dei valori semantici delle proposizioni. Supporre un meccanismo in grado di capire i significati dei simboli usati dalla mente sarebbe stato come collocare un piccolo interprete (od omuncolo) nella testa. Dall'altro lato è difficile vedere come un processo specificato in termini puramente causali possa contare come processo di ragionamento, visto che chiamare qualcosa «ragionamento» significa collocarlo rispetto a delle norme e non solo rispetto a delle cause (cioè chiamare un processo «razionale» non è solo descrivere la sua eziologia causale, ma significa dire che rispetta, o almeno che è valutabile secondo alcuni standard di ragionamento, come ad esempio la validità).

La TCM (o, più specificamente, il CAR) può essere vista come una prova di compatibilità, che mostra la conciliabilità del realismo intenzionale (ossia una responsabilità rispetto alla realtà delle proprietà semantiche degli stati mentali, e ai ruoli causali di stati mentali nella determinazione del comportamento) con l'affermazione che tali processi mentali sono tutti processi causali per i quali un meccanismo causale potrebbe, in principio, essere specificato. Il trucco di collegare la semantica alla causalità è di unirli in posizione intermedia alla sintassi. La formalizzazione ci mostra come collegare la semantica alla sintassi, mentre la computazione ci mostra come connettere la sintassi ai meccanismi causali. Perciò esiste un modello consistente sul quale i processi di ragionamento *bona fide* (processi che rispettano il valore semantico dei termini) possono essere condotti attraverso meccanismi fisici non più misteriosi: ciò significa che se la mente è un computer nel senso che le sue rappresentazioni mentali sono tali per cui tutte le proprietà semantiche sono condotte collegando le proprietà sintattiche che possono essere utilizzate da una «macchina sintattica» (Haugeland, 1981), tutto ciò è causalmente responsabile del ragionamento.

Una prova della compatibilità è di per sé una prova debole per la veridicità di una teoria. Comunque, negli anni '80 e '90, molti filosofi furono convinti dall'affermazione di Fodor che la TCM è «l'unico gioco in città» (cioè che le uniche conoscenze che abbiamo dei processi cognitivi sono computazionali e questo implica postulare un linguaggio del pensiero e delle operazioni condotte sulle rappresentazioni in questo linguaggio). Data l'argomentazione che la TCM è implicita nelle teorie delle scienze cognitive, la sua ulteriore abilità a fornire una prova di compatibilità per il fiscalismo ed il realismo intenzionale ha rafforzato le sue credenziali filosofiche mostrando che l'interpretazione delle scienze cognitive era produttiva anche dal punto di vista filosofico.

2.2 *La semantica degli stati mentali*

Oltre alla prova della compatibilità, alcuni filosofi hanno visto la TCM – o più precisamente la TRM – anche come una spiegazione delle proprietà semantiche degli stati mentali. Fodor, per esempio, afferma che come i discorsi pubblici ereditano le loro proprietà semantiche dai pensieri del parlante, i pensieri ereditano le loro proprietà semantiche dalle rappresentazioni mentali del LOT, che sono tra i loro costituenti. Se ho un pensiero che si riferisce a Bill Clinton, è perché questo pensiero è in relazione alla rappresentazione mentale che si riferisce a Bill Clinton. Se penso «Clinton è stato presidente nel 1995» è perché sono in una particolare relazione funzionale (caratteristica della credenza) che ha il contenuto «Clinton è stato presidente nel 1995».

2.3 *Legami con la scienza cognitiva e la IA*

Oltre a questi potenziali contributi alla filosofia della mente, la TCM era in relazione simbiotica con le applicazioni della visione della mente come computer nell'intelligenza artificiale e nelle scienze della cognizione. Da un lato, formulazioni filosofiche come la TCM articolavano una visione generale della mente e della computazione che era congeniale a molti

ricercatori in IA (Intelligenza Artificiale) e in scienza cognitiva. Dall'altro lato, i successi dei modelli computazionali del ragionamento, del linguaggio e della percezione hanno conferito credibilità all'idea che questo processo possa essere compiuto anche mediante la computazione della mente. Due legami con la ricerca empirica sono evidenti per la loro importanza storica. Il primo legame è con Chomsky, il quale introdusse una «rivoluzione cognitiva» in campo linguistico che soppiantò l'allora prevalente visione comportamentista dell'acquisizione linguistica. Quest'ultima, secondo Chomsky (1959), non era in grado di spiegare il fatto che un bambino capisca le regole grammaticali e sia poi capace di applicarle in un numero indefinito di nuovi contesti in modi determinati dal limitato insieme di stimoli ai quali è stato esposto. Ciò, secondo Chomsky, richiedeva il postulato di un meccanismo che non funzionasse semplicemente su principi generali di condizionamento classico e operante, ma fosse specificamente ottimizzato per l'acquisizione linguistica. La linguistica chomskyana parlava spesso degli sforzi del bambino per avere padronanza della grammatica in termini di formazione e rafforzamento di ipotesi; e questo, sosteneva Fodor (1975), richiedeva un linguaggio di pensiero interiore. Si riteneva perciò che la linguistica chomskyana avesse bisogno della TRM, e i computazionalisti hanno ritenuto plausibile che i meccanismi basilari potessero essere eseguiti in termini computazionali.

La grammatica chomskyana sottolineava inoltre le caratteristiche dell'abilità linguistica come la *sistematicità* (la persona in grado di capire la frase «il cane inseguiva il gatto» può capire anche la frase «il gatto inseguiva il cane») e *produttività* (l'abilità di una persona di possedere una quantità infinita di pensieri generati da un numero limitato di basi lessicali e regole sintattiche ricorsive). Anche il pensiero, ovviamente, ha le stesse caratteristiche. Da una prospettiva computazionalista, questi due fatti non sono collegati casualmente: il linguaggio naturale è sistematico e produttivo perché è un'espressione dei pensieri di una mente che già possiede sistematicità e produttività; la mente a sua volta possiede que-

ste caratteristiche perché il pensiero si manifesta come un sistema rappresentazionale sintatticamente strutturato. In realtà, afferma Fodor, una lingua strutturata sintatticamente è l'unica maniera conosciuta per garantire queste caratteristiche; perciò, fino a prova contraria, si può credere che la TRM sia attendibile.

Il secondo importante legame è tra la scienza cognitiva e la teoria della visione di David Marr (Marr, 1982; Marr e Poggio, 1977), il primo a sperimentare un approccio computazionale alla visione. Mentre pochi dei dettagli presi da loro in considerazione sono sopravvissuti all'attuale lavoro nella scienza della visione, ciò che era più influente nella loro opera non erano i dettagli empirici, ma una serie di importanti idee metateoretiche. Secondo Marr in una teoria della visione è necessario distinguere tre livelli (e, per estensione, è necessario farlo anche per altri processi cognitivi). Al livello più alto c'era una specificazione del compito che deve svolgere un sistema: per esempio, nel caso della visione, quello di costruire una rappresentazione tridimensionale degli stimoli distali sulla base degli input giunti alla retina. Ciò è definito da Marr (piuttosto infelicemente) «livello computazionale». All'altro capo dello spettro c'era un livello che descriveva «l'adempimento» di questa funzione da parte dell'«hardware» del sistema (per esempio, le proprietà neurochimiche alla base della fototrasduzione nelle cellule della retina). Questi due livelli in sé rappresentano un quadro funzionalista convenzionale. Tra i due livelli Marr ha inserito un livello «algoritmico» di spiegazione. Il compito del teorico era di isolare un candidato per l'algoritmo utilizzato dal sistema al fine di svolgere il lavoro. Un algoritmo appropriato al compito specifico a livello «computazionale», e compatibile con i fatti neurologici a livello «implementazionale».

Tale livello algoritmico intermedio è ovviamente in stretta connessione con una strategia per i processi di modellamento visuale: il modellatore parte da dati psicofisici (leggi di Weber) per poi provare a costruire dei modelli con condizioni di input/output isomorfo. Gli algoritmi per la trasformazione degli input in output sono utilizzati come modelli computa-

zionali. Mentre è possibile vedere questo modellamento alla pari con il modellamento di sistemi meteorologici (nei quali non si suppone che ciò che è stato modellato sia in qualche modo interessante dal punto di vista «algoritmico» o «computazionale») la disponibilità di tecniche di modellamento computazionale suggerisce l'ipotesi secondo cui i processi visivi sono compiuti algebricamente, dato che i metodi algoritmici fanno per lo meno parte dei metodi per portare a termine i compiti informativi implicati nei dati psicofisici. C'è ovviamente una nota ambiguità filosofica latente che riguarda la confusione nel comportarsi in una maniera *descrivibile* da una regola – con ciò che ne segue – oppure *applicando* una regola: discutibilmente, i sostenitori di un livello algoritmico di descrizione non hanno sempre tenuto a mente questa distinzione. Ciononostante, con questo avvertimento (cioè di non aver chiaro se il livello algoritmico sia semplicemente un livello di descrizione o se sia previsto che il sistema debba applicare un algoritmo), alcune versioni dell'approccio a tre livelli di Marr sono diventate ben presto qualcosa di ortodosso nella scienza cognitiva negli anni '80.

L'approccio di Marr è ovviamente collegato alla TCM. Entrambi prevedono rappresentazioni interne e processi algoritmici che mediano le trasformazioni da una rappresentazione a un'altra. L'aumento di modelli che utilizzano l'approccio a tre livelli di Marr nelle scienze della cognizione sembrava fornire un supporto empirico alla visione della mente come un algoritmico processore di simboli. Inoltre la ricerca ha proposto una morale per la TRM e la TCM, rendendole, in un certo senso, potenzialmente più radicali. La TRM e la TCM avrebbero potuto essere valide solo come teorie dei tipi di processi mentali che possono essere articolati in frasi in un linguaggio naturale – processi di «alto livello» come i pensieri consci. Comunque gli algoritmi di Marr si applicano ad un livello molto più semplice, come i processi di informazione tra due livelli di cellule nel sistema visivo. Questi processi non sono soggetti a verifiche o interventi consci, e una persona non sarebbe in grado di esprimerli a parole. Pertanto tale teoria implica il postulato di un insieme di simboli ed algoritmi,

non del tutto inconsci, ma infraconsci – ossia, processi che hanno luogo in un livello di gran lunga più semplice di ciò che i filosofi sono abituati a considerare «pensieri».

3. *Le critiche alla teoria*

3.1 *La sintassi spiega la semantica?*

Come già detto, la relazione più forte proposta tra la sintassi e la semantica dei simboli – le proprietà semantiche sopravvivono su quelle sintattiche – non fu mai abbracciata dai maggiori fautori della TCM. Putnam (1980), ad esempio, aveva mostrato un ulteriore ostacolo alla teoria, che consiste in una conseguenza del teorema della logica di Lowenheim-Skolem, dal quale ne consegue che ogni sistema di simbolo formale ha almeno un'interpretazione nella teoria del numero. In questo senso, prendiamo ogni descrizione sintattica D di «mentalese». Dato che i nostri pensieri non sono solo numeri, un'interpretazione canonica della semantica del «mentalese» (che chiameremo S) avrebbe bisogno di collegare almeno alcuni dei termini di riferimento ad oggetti non matematici. Ad ogni modo, Lowenheim-Skolem assicura che esiste almeno un'interpretazione S^* che collega tutti i termini di riferimento ad oggetti puramente matematici. S^* non può essere l'interpretazione canonica, ma non c'è nulla nella sintassi del «mentalese» che sia in grado di spiegare perché S è l'interpretazione corretta, mentre S^* non lo è. Pertanto la semantica è determinata dalla sintassi (Cfr. Pylyshyn, 1984: 44).

3.2 *Il vocabolario semantico è applicato univocamente?*

Fodor (1981) afferma che la teoria secondo la quale gli stati mentali sono in relazione con le rappresentazioni simboliche dovrebbe spiegare come gli stati mentali abbiano valori semantici ed intenzionalità. Questo perché, secondo Fodor, sono le rappresentazioni mentali ad avere queste proprietà «in prima istanza», mentre gli stati di attitudine proposiziona-

le le «ereditano» dalle rappresentazioni mentali che sono tra i loro costituenti. Questa teoria inizialmente è stata criticata da Searle (1980, 1984) e Sayre (1986, 1987), per poi essere approfondita da Horst (1996). La critica è stata brevemente riassunta come segue. Supponiamo di rappresentare schematicamente l'affermazione di Fodor così:

– F: uno stato mentale M corrisponde a P perché la rappresentazione mentale MR corrisponde a P.

Una simile affermazione è più plausibile con il presupposto che l'espressione «... corrisponde a P» è univoca nelle due accezioni in F – cioè, che «... significa P» funziona allo stesso modo quando è applicata a stati mentali (come credenze, desideri e giudizi) e a rappresentazioni mentali (cioè, simboli in un linguaggio di pensiero). Con questo presupposto, il «significato» delle rappresentazioni mentali è potenzialmente atto a spiegare il «significato» degli stati mentali, perché è esattamente la stessa proprietà di «significato» in questione in entrambi i casi.

Tuttavia, questo presupposto è in tensione con quello secondo cui il tipo di «significato» attribuito alle rappresentazioni mentali è lo stesso tipo di «significato» attribuito a simboli come discorsi e iscrizioni. Secondo i critici, attribuzioni di significato, come «questa iscrizione significava P» hanno una celata complessità nella loro struttura logica. Il verbo «significa» non esprime semplicemente una doppia relazione tra iscrizione e il suo valore semantico; piuttosto, deve velatamente riportare o (a) quel che intende l'oratore, o (b) l'interpretazione di chi ascolta, o (c) l'interpretazione fornita da una specifica convenzione linguistica. Trattandosi di un significato specificamente simbolico («significato» nel senso di parola applicata a simboli senza che siano previsti altri usi del termine) che la formalizzazione e la computazione ci mostrano come collegare alla sintassi, è importante che sia questo il suo uso nell'ambito della TCM. Ma se supponiamo che la frase «MR significa P» in F deve essere eseguita secondo ciò che intende l'oratore, l'interpretazione di chi ascolta, o

l'interpretabilità convenzionale, non ci troveremmo in nessuno di questi casi con una potenziale spiegazione del tipo di «significato» attribuito agli stati mentali. Ognuna di queste nozioni è infatti concettualmente dipendente dalla nozione di stati mentali significativi, e quindi il significato mentale non può essere spiegato in termini di significato simbolico senza essere coinvolto in un circolo esplicativo.

Ci sono stati due tipi di risposte esplicite a questa critica, e probabilmente anche una terza molto più implicita. Secondo una di queste repliche esiste un uso univoco del vocabolario semantico: vale a dire, ciò che affermano le teorie semantiche come quella di Tarski, secondo cui «una semantica» è una semplice mappatura dai tipi di simbolo alle loro estensioni, o un'efficace procedura per produrre questa mappatura. Comunque, questa visione della «semantica» è probabilmente troppo debole per essere esplicativa, dato che esiste un numero indefinito di rapporti di mappatura, e solo pochi di questi sono anche semantici (cfr. Blackburn, 1984; Field, 1972; Horst, 1996). Una seconda replica è guardare ad una semantica alternativa «più solida», come quella di C. S. Peirce (cfr. von Eckardt, 1993). Probabilmente non ci sono stati sufficienti esami del rapporto tra la semantica di Peirce e il punto di vista equivoco che differenzia «significato mentale» (vale a dire: il senso di «significato» può essere applicato agli stati mentali) da significato di chi parla e dall'interpretabilità convenzionale dei simboli. In un certo senso la strategia di Peirce si collega ad una terza replica mossa dai computazionalisti, che potrebbe anche segnare un punto d'incontro tra le precedenti.

I computazionalisti hanno di norma approvato le teorie causali della semantica delle rappresentazioni mentali. Nonostante il proprio punto di vista sulle prospettive generali delle teorie causali del significato, un senso del «significato» inteso in termini di covarianza causale o eziologia causale non può equivalere né al significato di chi parla né all'interpretabilità convenzionale. La buona notizia per il computazionalista è che ciò può salvare la teoria dalla circolarità esplicativa e dal regresso. In ogni modo, è probabile che metta a rischio la

prova di compatibilità della TCM: il tipo di «significato» in grado di essere unito alla sintassi nelle macchine computatrici è quello convenzionale: per esempio, tale sequenza di cifre binarie è interpretabile secondo alcune convenzioni come un particolare numero intero. Ma se il tipo di «significato» attribuito alle rappresentazioni mentali non è di questo tipo, allora gli elaboratori non hanno dimostrato che tali tipi di «significato» possono essere legati alla sintassi nei modi opportuni. Sebbene questo problema non sia insormontabile, può essere ancora visto come una questione aperta. Allo stesso modo, l'attuabilità di tale strategia è ulteriormente dipendente dalle possibilità della semantica causale di spiegare il tipo di «significato» attribuito agli stati mentali (cfr. Horst, 1996).

3.3 Sono le nostre abilità cognitive formalizzabili e computabili?

Alcuni recenti critici della CTM sono partiti dall'osservazione che non tutti i processi sono computabili (cioè riducibili ad una soluzione algoritmica), e hanno concluso che le spiegazioni computazionali sono possibili soltanto per quei tipi di processi mentali riducibili a tecniche algoritmiche. Tuttavia, c'è una valida ragione di credere che esistono campi problematici sui quali gli esseri umani possono riflettere e su cui è possibile avere qualche conoscenza ma che non sono formalmente computabili.

La più antica linea argomentativa è dovuta a J.R. Lucas (1961), che ha desunto in una serie di articoli che il teorema dell'incompletezza di Gödel pone problemi alla tesi secondo cui la mente funziona come un computer. Più recentemente, Penrose (1989, 1990), ha sviluppato argomenti che portano alla medesima conclusione. La tesi di fondo di questi argomenti è che i matematici umani, di fatto, conoscono l'aritmetica molto meglio di un computer. Pertanto ci deve essere qualcosa in questo tipo di cognizione umana della semplice computazione. Si è sviluppato un esteso dibattito su questo punto negli ultimi quarant'anni (vedi ad esempio le critiche in Bowie, 1982; Feferman, 1996), e la continua di-

scussione suggerisce che, in questo genere di argomenti, la questione è ancora apertissima.

Una diversa linea di argomenti è stata sviluppata da Hubert Dreyfus (1972). Questi sostiene che la maggior parte delle competenze e conoscenze umane non può di fatto essere ridotta ad una procedura algoritmica, e che pertanto non sia computabile in un senso tecnico rilevante. Soffermendosi sulle concezioni di Heidegger e della fenomenologia esistenziale, Dreyfus punta alla differenza principale tra il tipo di cognizione che si può impiegare quando si acquisisce una competenza tecnica e il tipo impiegato dall'esperto. Lo scacchista principiante può seguire regole come «alla prima mossa avanza la pedina del Re di due caselle» e così via. Ma seguire tali regole è precisamente il segno del principiante. Il maestro di scacchi viceversa semplicemente «vede» la «mossa giusta». Ciò significa non essere impegnato nel seguire regole ma semplicemente nel concretizzare una attività specializzata. Dreyfus illustra le sue considerazioni riferendosi al problema posto dai ricercatori dell'IA che hanno tentato di codificare la conoscenza esperta in programmi di computer. Il successo o il fallimento hanno qui realmente poco a che fare con i meccanismi di computazione: il problema è infatti se le competenze esperte nel campo in questione possano essere catturate da una procedura algoritmica. In certi campi ben circoscritti ciò è potuto accadere; ma più spesso appare impossibile, argomenta Dreyfus, «catturare» conoscenze esperte in un algoritmo, in modo particolare laddove sono richieste conoscenze generali che oltrepassano il campo problematico.

Ci sono due tipi di risposte possibili alle critiche di Dreyfus. La prima è che Dreyfus conferisce troppo peso allo stato attuale, iniziale, delle ricerche di IA, e trarrebbe delle inferenze circa ogni possibile sistema basato su regole solo sulla base degli insuccessi di particolari esempi di una tecnologia evidentemente ancora agli inizi. Una parte di queste critiche è sicuramente corretta: gli argomenti di tipo induttivo di Dreyfus rischiano di essere sconfessati da ricerche future. Egli potrebbe però replicare che l'ottimismo dei ricercatori dell'IA è parimenti criticabile, ma dobbiamo concludere che una ge-

neralizzazione induttiva pura su ciò che i computer non possono fare si rivela un argomento piuttosto debole. In ogni caso, l'argomento di Dreyfus non è solo puramente induttivo; esso contiene una più decisiva posizione circa la natura delle performance esperte e circa la non praticabilità di tecniche basate su regole atte a duplicare questo tipo di performance. Si tratta di una questione complessa, che non ha ancora ricevuto né un'accoglienza decisiva né un rifiuto netto tra i filosofi della mente.

Il secondo tipo di risposte a Dreyfus implica concedere che vi possono essere dei problemi (forse anche problemi gravi) per certi tipi di sistemi – come quelli basati su regole – ma che esistono tuttavia altri sistemi in grado di evitarli. Si può pensare alle strategie «bottom-up» delle reti connessioniste o al tentativo di Rodney Brooks di costruire livelli di intelligenza semplice tipo quella degli insetti all'interno di un approccio assai promettente rispetto alle critiche che Dreyfus effettua alle limitazioni dei sistemi basati su regole. Dreyfus stesso sembra aver sperimentato entrambe le facce della sua posizione. In un articolo del 1988 redatto assieme a suo fratello Stuart, egli si mostra interessato al fatto che le reti neurali assumono oggi un ruolo di rilievo su queste tematiche, e si mostra naturalmente interessato anche ai campi problematici nei quali gli approcci basati su regole e rappresentazioni hanno incontrato problemi. In ogni caso, nell'edizione del 1992 di *What Computers Still Can't Do* egli afferma che «l'approccio connessionista prima ripudiato poi recuperato sta semplicemente ottenendo la sua meritevole chance di fallire». In un senso più stretto, si può comunque rilevare che questa serie di obiezioni cede a Dreyfus su un punto effettivo, secondo cui non è che non possa esistere una parte di hardware intelligente chiamato «computer», quanto piuttosto che non si può costruire intelligenza o cognizione fuori della «computazione» nel senso della «manipolazione di simboli basati su regole».

3.4 *La computazione è sufficiente per la comprensione?*

Forse la critica più autorevole alla CTM rimane l'esperimento immaginario di John Searle (1980) noto come «stanza cinese». In questo esperimento ad un essere umano viene conferita la posizione della CPU in un computer. Egli viene collocato in una stanza senza alcuna possibilità di comunicare con l'esterno eccetto che per tracce simboliche immesse nella stanza attraverso una fessura. Si tratta di simboli, dotati di significato, ma scritti in cinese, una lingua che egli non conosce. L'uomo tenta di fornire risposte significative e appropriate e le fa passare al di fuori della stanza. In questo lavoro può contare sull'aiuto di un testo di grammatica che contiene le regole utilizzabili per decidere quali simboli occorre scrivere in risposta a particolari condizioni di input. Questo esperimento riproduce il funzionamento di un computer digitale: anche esso infatti può ricevere e produrre input simbolici; anche esso può manipolare i simboli che riceve sulla base di regole totalmente corrispondenti a quelle che potrebbero essere applicate sulla base di informazioni non semantiche come la sintassi e la pura forma simbolica. La sola differenza nell'esperimento immaginario è che in esso l'«unità di processo» è affidata ad un essere umano.

Questo esperimento fornisce una replica diretta al suggerimento di Alan Turing di sostituire la domanda «le macchine possono pensare?» con la domanda se esse possono subentrare in un «gioco di imitazione» nel quale a coloro che pongono domande viene richiesto di determinare, esclusivamente sulla base delle risposte ricevute, se un interlocutore non visibile sia una persona oppure una macchina. Questo, come noto, è stato chiamato «Turing Test». La risposta di Searle, nella sostanza, è questa: assumiamo che sia possibile costruire un programma – un insieme di regole – che permetta ad una macchina di utilizzare regole che le consentano di superare il test di Turing. Ebbene, la capacità di superare il test è davvero sufficiente a stabilire che la «cosa» che ha superato il test è realmente una «cosa che pensa»? La stanza cinese di Searle possiede le caratteristiche di una macchina in grado di supe-

rare il Turing Test in cinese: produce risposte significative e appropriate dal punto di vista della conversazione e riesce a fare ciò utilizzando esclusivamente valori sintattici. Ma quando ci domandiamo se la stanza cinese, o una parte di essa (la persona che vi sta dentro) può comprendere le frasi che le vengono sottoposte la risposta è naturalmente negativa. Infatti, la persona all'interno della stanza *non* conosce la lingua cinese, e né un libro di regole né un sistema uomo/libro-di-regole/stanza sembrano essere in grado di stimolargli una qualche comprensione. Il risultato è chiaro: se anche fosse possibile simulare una competenza linguistica ricorrendo a puri significati computazionali, ciò non costituirebbe comunque una condizione *sufficiente* affinché si abbia comprensione autentica.

L'argomento della stanza cinese ha goduto di una longevità lunga quasi quanto quella della stessa CTM. Esso ha prodotto una piccola industria di articoli filosofici, le tesi di molti dei quali erano già state anticipate nei commenti pubblicati in *Behavioral and Brain Sciences* nel 1980. Alcuni critici hanno sostenuto che la «comprensione» può essere definita in termini integralmente funzionali, e che quindi anche il sistema della stanza cinese esibisce in realtà una forma di comprensione. Altri hanno concesso che il congegno descritto da Searle non mostra certo una comprensione, ma hanno tuttavia ipotizzato che essa potrebbe essere raggiunta con l'aggiunta di ulteriori caratteristiche – un corpo da robot, un apparato sensoriale, la capacità di apprendere nuove regole e informazioni, l'immersione in un ambiente reale con il quale interagire. Searle ed altri autori hanno adattato l'esperimento immaginario in modo tale da prevedere questo tipo di variazioni. Tali adattamenti, come l'esperimento originale, sembrano però stimolare costantemente intuizioni diverse in lettori diversi: la stanza cinese è pertanto rimasta, nella letteratura su questi argomenti, uno dei contributi più inquietanti e provocatori per il pensiero.

3.5 La CTM è il «solo gioco in città»?

Una pietra angolare per la CTM è stata il fatto che alcune versioni della teoria sono state legate, nelle teorie cognitive, ai processi di apprendimento e acquisizione del linguaggio, e che tali teorie sono state a lungo le uniche concorrenti in questi campi. I critici della CTM hanno sostenuto però che attualmente, per la maggior parte dei fenomeni psicologici, esistono spiegazioni alternative che non richiedono ragionamenti basati su regole in un linguaggio del pensiero. Negli anni '80, alcuni filosofi hanno infatti cominciato a prendere coscienza del paradigma dei processi di *modeling* psicologico, definito talvolta anche come approccio delle «reti neurali» o «connessionista». Tale approccio è stato sviluppato formalmente e messo alla prova da un punto di vista empirico inizialmente dall'opera di Wiener e Rosenblatt, diffondendosi in seguito, tra gli anni '60 e oggi, grazie a ricercatori come Grossberg e Anderson. Si trattava all'inizio soltanto di ricognizioni filosofiche sulle prime ricerche cibernetiche (cfr. Sayre, 1969, 1976); i modelli di reti neurali sono entrati infatti nel grande dibattito filosofico soltanto dopo la pubblicazione di *Parallel Distributed Processing* (1986) di Rumelhart e McClelland.

I modelli di reti neurali mirano a riprodurre la dinamica dei processi psicologici, non direttamente a livello degli stati intenzionali ma a quello delle reti di neuroni attraverso cui (presumibilmente) gli stati mentali si realizzano. In qualche caso, fenomeni psicologici che si sono mostrati resistenti alla modellizzazione algoritmica a livello cognitivo sembrano rientrare soltanto nell'architettura delle reti neurali, oppure modelli di reti di particolare disegno. Alcuni tipi di apprendimento sembrano essere ricompresi naturalmente nelle architetture di rete, e più recentemente ricercatori come Smolensky hanno prodotto risultati che suggeriscono come alla fine alcune caratteristiche dell'acquisizione del linguaggio possano essere perfettamente simulate da questi modelli.

Tra il 1980 e il 1990 vi sono state molte discussioni filosofiche sulla relazione tra modelli di reti e modelli computazionali della mente. Le architetture connessioniste sono state

contrastate dalle architetture classiche o «GOFAI» («good old-fashioned AI», cara vecchia intelligenza artificiale), le quali impiegano regole e rappresentazioni simboliche. Avvocati difensori del connessionismo, come Smolensky (1987), hanno sostenuto che i modelli connessionisti si distinguono in maniera rilevante dai classici modelli computazionali sulla base del fatto che in essi il processo implicato avviene ad un sottolivello simbolico. Diversamente che in un computer convenzionale, il processo è infatti distribuito più che seriale, non vi è una esplicita rappresentazione delle regole e le rappresentazioni non sono concatenate.

Vi è però generale accordo circa il fatto che alcune di queste differenze non sono rilevanti. Entrambi i punti di vista concordano, ad esempio, sul fatto che i processi cerebrali sono paralleli e distribuiti. Similmente, anche nei modelli di produzione dei computer, soltanto in programmi di riserva le regole vengono rappresentate: le regole «hard-wired» nella CPU, ad esempio, non lo sono.

La più importante risposta «classica» è quella di Fodor e Pylyshyn (1988, ma vedi anche Fodor e McLaughlin, 1990). Essi hanno sostenuto che alcuni sistemi connessionisti in grado di garantire sistematicità e produttività potrebbero essere semplicemente una concretizzazione (*implementation*) dell'architettura classica (LOT). C'è molta disparità, in ogni caso, su quali siano esattamente le caratteristiche costitutive di una architettura classica. Van Gelder (1991), ad esempio, ritiene che i classicisti facciano riferimento ad una specifica «composizionalità concatenativa» (Van Gelder, 1991: 365). Ciò significa che le caratteristiche cognitive possono non essere puramente «implementazionali», segnando così una significativa differenza rispetto al classicismo. In risposta, Aydede (1997), mentre effettua una ricognizione della tendenza dei classicisti ad assumere che la LOT è concatenativa, desume che essa avrebbe bisogno di non essere ristretta all'interno di criteri forti (cfr. Loewer e Rey, 1991). In ogni caso, se si concede che i sistemi non concatenativi come quello di Smolensky o il «Recursive Auto-Associative Memory» di Pollack costituiscano esempi di concretizzazione del

LOT, più attenzione andrebbe prestata a come la nozione di «linguaggio» del pensiero implichi delle difficoltà su quale tipo di «rappresentazioni» siano incluse o escluse dalla famiglia dei modelli LOT. Non ci sono accordi generali su come risolvere queste dispute particolari: sulla base di un notevole flusso di articoli, la questione può essere classificata come «aperta».

Riguardo alla CTM, la ricognizione o i modelli alternativi di reti (e altri alternativi modelli, come l'approccio ai sistemi dinamici di van Gelder) hanno alla fine colpito l'argomento dell'«unico gioco in città». Nell'attuale situazione dialettica, i difensori del CTM debbono chiarificare ulteriormente la relazione esistente tra i loro modelli e i modelli neurali, e debbono attestare come i loro modelli costituiscano descrizioni migliori delle cose che avvengono nella mente e nel cervello umani in alcuni campi problematici.

3.6 *La psicologia grida vendetta?*

Le critiche sin qui riportate non minacciano evidentemente le intenzioni della CTM di presentare una prova della compatibilità tra il realismo intenzionale e il fiscalismo, almeno per quanto riguarda i casi di comprensione formalizzabili. Questo scopo – la psicologia che si «vendica» dimostrando la propria compatibilità con la fisica generale – ha rappresentato esso stesso una parte importante del movimento computazionale in psicologia e ciò spiega perché le teorie rappresentazionali/computazionali siano state spesso viste come le alternative principali al materialismo eliminativista tra il 1980 e il 1990.

In ogni modo, questo stesso scopo, ed il corrispondente impegno per un particolare tipo di naturalizzazione della mente, è insufficientemente soggetto ad analisi nello scenario attuale. Poniamo dunque la questione seguente: se lo scopo era intensificare l'impegno dell'uno in favore dei risultati di una qualche scienza empirica come la psicologia e l'impegno dell'altro per una posizione metafisica (come il materialismo) o una metateoria sulla scienza (come l'ipotesi dell'Unità delle

Scienze), quale altra via resta da percorrere? È curioso, in un certo senso, che Fodor, un importante difensore dell'autonomia delle scienze speciali, tra cui in particolare la psicologia (cfr. Fodor, 1974), mostri di prorogare l'impegno metafisico o metateorico vedendo la psicologia come una scienza che ha bisogno di vendetta. Al contrario, i filosofi della scienza, dal 1970, sono stati fortemente spinti a rifiutare alcuni standards metateorici e metafisici, e più specificamente hanno lavorato in favore dell'autonomia delle scienze «locali», con l'assunzione che le scienze debbono uniformarsi solo in qualche caso particolare. (In questo spirito, l'originale inventore del CTM, Hilary Putnam, ha recentemente abbracciato un pluralismo pragmatistico). È dunque possibile che in tempi futuri la recente preoccupazione della vendetta della psicologia potrà essere vista come una delle ultime vestigia del movimento positivista dell'Unità della Scienza.

Bibliografia

- AYDEDE, M., *Language of Thought: A Connectionist Contribution*, in «Minds and Machines» 7, 1997, pp. 57-101.
- BLACKBURN, S., *Spreading the Word: Groundings in the Philosophy of Language*. New York, Oxford University Press, 1984.
- BODEN, M., *The Philosophy of Artificial Intelligence*. Oxford University Press, 1990.
- BOWIE, L., *Luca's Number is Finally Up*, in «Journal of Philosophical Logic» 11, 1982, pp. 279-85.
- CHOMSKY, N., *A Review of B.F. Skinner's Verbal Behavior*, in «Language», 35, 1959, pp. 26-58.
- CUMMINS, R., *Meaning and Mental Representation*, Cambridge, Mass., MIT Press, 1989.
- DREYFUS, H., *What Computers Can't Do*, New York, Harper and Row, 1972.
- , *What Computers Still Can't Do*, Cambridge, Mass., MIT Press, 1991.

- DREYFUS, H., STUART D., (1988), *Making a mind versus modelling a brain: artificial intelligence back at a branch-point*, in BODEN 1990, pp. 309-333.
- FEFERMAN, S. *Penrose's Goedelian Argument*, in «Psyche» 2, 1996, pp. 21-32.
- FIELD, H., *Tarski's Theory of Truth*, in «Journal of Philosophy», 59, 1972.
- FODOR, J., *Special Sciences, or Disunity of Science as a Working Hypothesis*, in «Synthese» 28, 1974 pp. 97-115.
- , *The Language of Thought*, New York, Thomas Crowell, 1975.
- , *Methodological Solipsism Considered as a Research Strategy in Cognitive Science*, in «Behavioral and Brain Sciences», 3, 1980, pp. 63-73.
- , *Representations*, Cambridge, Mass., Bradford Books/MIT Press, 1981.
- , *Psychosemantics*. Cambridge, Mass., Bradford Books, 1987.
- , *A Theory of Content and Other Essays*. Cambridge, Mass., Bradford Books, 1990
- , *The Elm and the Expert*. Cambridge, Mass., Bradford Books, 1993.
- , *The Mind Doesn't Work That Way*. MIT Press, 2000.
- FODOR, J., MCLAUGHLIN, B., *Connectionism and the Problem of Systematicity: Why Smolensky's Solution Doesn't Work*, in «Cognition» 35, 1990, pp. 193-204.
- HAUGELAND, J., *The Nature and Plausibility of Cognitivism*, in «Behavioral and Brain Sciences», 2, 1978, pp. 215-226.
- (a cura di), *Mind Design*. Cambridge, Mass., MIT Press/Bradford Books, 1981.
- HORST, S., *Eliminativism and the Ambiguity of 'Belief'*, in «Synthese», 104, 1995, pp. 123-145.
- , *Symbols, Computation and Intentionality: A Critique of the Computational Theory of Mind*. Berkeley and Los Angeles, University of California Press, 1996.
- , *Symbols and Computation*, in «Minds and Machines», Volume 9, No. 3, 1999, pp. 347-381.
- LEWIS, D., *Lucas Against Mechanism*, in «Philosophy», 44, 1969, pp. 231-233.

- , *Lucas Against Mechanism II*, in «Canadian Journal of Philosophy», 9, 1979, pp. 373-376.
- LUCAS, J.R., *Minds, Machines, and Godel*, in «Philosophy», 36, 1961, pp. 112-127.
- MARR, D., *Vision: A computational Investigation into the Human Representation and Processing of Visual Information*, New York, W. H. Freeman and Company, 1983.
- MARR, D., T. POGGIO., *From understanding computation to understanding neural circuitry*, in «Neurosciences Res. Prog. Bull.», 15, 1977, pp. 470-488.
- PENROSE, R., *The Emperor's New Mind*. Oxford University Press, 1989.
- , *Précis of The Emperor's New Mind*, in «Behavioral and Brain Sciences» 13, 1990, pp. 643-705.
- PUTNAM, H., *Minds and Machines*, in «Dimensions of Mind», a cura di S. Hook, New York, New York University Press, 1960.
- , *Brains and Behavior*, originariamente parte del programma dell'«American Association for the Advancement of Science», Sezione L («History and Philosophy of Science»), Dicembre 27, 1961. Ristampato in BLOCK 1980.
- , *The Nature of Mental States*, in «Art, Mind and Religion». A cura di W.H. CAPITAN, D.D. MERRILL, Pittsburgh, University of Pittsburgh Press. Ristampato in BLOCK 1980.
- , *Models and Reality*, in «Journal of Symbolic Logic», 80, 45, pp. 464-482.
- PYLYSHYN, Z., 1980, *Computation and Cognition: Issues in the Foundation of Cognitive Science*, in «The Behavioral and Brain Sciences» 3, 1980, pp. 111-132.
- , *Computation and Cognition: Toward a Foundation for Cognitive Science*, Cambridge, Mass, Bradford Books/MIT Press, 1984.
- RUMELHART, D., E., J., MCCLELLAND, in collaborazione con PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, Mass., MIT Press, 1986.
- SAYRE, K., *Consciousness: A Philosophic Study of Minds and Machines*, New York, Random House, 1969.

Steven Horst

- , *Cybernetics and the Philosophy of Mind*, Atlantic Highlands, New Jersey, Routledge & Kegan Paul, 1976.
- , 1986, *Intentionality and Information Processing: An Alternative Model for Cognitive Science* in «Behavioral and Brain Sciences», 9(1), 1986, pp. 121-138.
- , in «Synthese», 70, pp. 247-269.
- SEARLE, J., *Minds, Brains and Programs*, in «Behavioral and Brain Sciences», 3, 1980, pp. 417-424.
- , *Minds, Brains and Science*, Cambridge, Mass., Harvard University Press, 1984.
- , Presidential Address, *Proceedings of the American Philosophical Association*, 1990.
- , *The Rediscovery of the Mind*, Cambridge, Mass., MIT Press, 1992.
- SMOLENSKY, P., *On the Proper Treatment of Connectionism*, in «Behavioral and Brain Sciences», 11, 1988, pp. 1-74.
- TURING, A., (1936), *On computable numbers*, in «Proceedings of the London Mathematical Society», 1936
- VON ECKARDT, B., *What is Cognitive Science?* Cambridge, Massachusetts, MIT Press, 1993.
- WINOGRAD, T., FLORES, F., *Understanding Computers and Cognition*, Norwood, New Jersey, Ablex Publishing Corporation, 1986.

FOLK PSYCHOLOGY⁵²
come simulazione mentale

Robert M. Gordon

La teoria della simulazione o «simulazione mentale» è la teoria della competenza psicologica umana del quotidiano, ossia delle abilità e risorse cui si fa comunemente appello per anticipare, spiegare e coordinare socialmente il comportamento. Secondo tale teoria, rappresentiamo le attività e i processi mentali degli altri tramite la loro simulazione mentale o la riproduzione di attività e processi simili in noi stessi: ad esempio, anticipiamo la soluzione altrui ad un problema di carattere pratico o teoretico risolvendo il problema noi stessi (adeguando le ovvie differenze). L'idea di base è che se le risorse di cui il nostro cervello fa uso per guidare il nostro comportamento sono passibili di modificazioni volte alla rappresentazione di altre persone, allora non v'è per noi alcun bisogno di compendiare esaustive informazioni riguardo le motivazioni del comportamento degli altri, in quanto gliele attribuiamo semplicemente. La simulazione è perciò definita processo-guida piuttosto che teoria-guida (Goldman, 1989).

La teoria della simulazione sembrò sfidare la supposizione in base alla quale una teoria (la psicologia del senso comune) sia alla radice del comportamento psicologico. Proprio tale aspetto suscitò grande interesse dal momento che senza questa supposizione la filosofia della mente si ritroverebbe senza fondamento: è questo, in altre parole, il dibattito esistente tra

⁵² R. M. Gordon, *Folks Psychology as Mental Simulation*, in «Stanford Encyclopedia of Philosophy», <http://plato.stanford.edu/entries/folkpsych-simulation/>

realisti psicologici, che hanno visto nella psicologia del senso comune una scoperta fondamentale valida per la scienza cognitiva, e gli eliminativisti, per cui invece si tratterebbe di una teoria fondamentale difettosa.

Diversamente dalle iniziali controversie concernenti il ruolo della conoscenza empatica e della ricostruzione storica nell'ambito delle scienze umane, il dibattito attuale tra la teoria della simulazione e la teoria della «teoria» si richiama a scoperte empiriche, tra cui i risultati sperimentali sulla competenza psicologica nello sviluppo dei bambini e le recenti scoperte nel settore della neuroscienza cognitiva.

1. Cosa si intende per «simulazione»?

Nei recenti dibattiti concernenti la psicologia del senso comune quotidiano, il termine «simulazione», come anche il termine «teoria», è stato usato ampiamente e nei modi più disparati. Solitamente la simulazione si ricollega all'assunzione di un ruolo o al processo immaginativo che ci consente di «metterci al posto di un altro». Tale metafora serve all'adozione di diverse prospettive spazio-temporali così come di altri espedienti in situazioni prototipiche (ad esempio, nelle interazioni sociali, d'ufficio o parentali); nonché dei tratti di un carattere alternativo e similari esercizi di personificazione drammatica. Tuttavia, la simulazione può anche essere vista come una semplice «proiezione», senza mutamenti immaginativi; non v'è quindi bisogno di mettersi al posto di un altro perché vi siamo già in tutto e per tutto: ad esempio, il tornado non si sta avvicinando a te o a me ma a noi.

Oltre a tale caratterizzazione della simulazione a livello individuale, essa è anche vista dalla maggior parte dei fautori in termini scientifico-cognitivi. Viene presupposto che, nell'assunzione del ruolo (*role-taking*), il sistema di controllo del comportamento individuale venga impiegato come modello manipolabile di altri sistemi simili. (Ciò non significa che la «persona» che simula sia il modello, ma piuttosto che solo il cervello individuale può essere manipolato per modellare altre persone). Il sistema è dapprima off-line, per cui le

emissioni non determinano il comportamento attuale ma solo predizioni o anticipazioni dello stesso e le immissioni e i parametri relativi al sistema non si limitano a quelli alla base del comportamento individuale. Quantunque ciò a volte sfoci nell'elaborazione di una decisione vicaria, più di frequente esso funge da deterrente per l'obiettivo più modesto di stabilire quali siano le opzioni attraenti (così si spiega il perseguimento d'una di esse) e quali no (e cioè invece alla base del rifiuto di perseguirle). Nell'ambito delle neuroscienze cognitive, il termine «simulazione» è usato per indicare l'attivazione (solitamente automatica ed inconsapevole), in risposta al comportamento osservato in un'altra persona, di meccanismi neuronali legati alla riproduzione individuale di tale comportamento.

Stephen Stich e Shaun Nichols, i cui saggi critici, chiarendo tali tematiche, sono stati fondamentali per la ridefinizione della teoria della simulazione, raccomandano l'eliminazione del termine «simulazione» per una nomenclatura più spicciola. Tuttavia, quanto mai importante risulta la precisione nella formulazione di ipotesi attendibili così come un'indagine ad ampio raggio e aperta sull'attuazione di procedimenti immaginativi top-down che filosofi e psicologi ritengono possano essere compatibili con i meccanismi *bottom-up* di recente scoperta.

2. Teoria della simulazione: tipologia

Alvin Goldman e lo psicologo Paul Harris dissentono da Robert Gordon e Jane Heal, i filosofi che, dopo studi indipendenti, hanno introdotto tale teoria nel 1986. (Il termine, coniato da Heal, «replicazione» e quello coniato da Gordon «simulazione» sono stati i comuni punti di riferimento del medesimo processo putativo). Per Gordon (1989) e, meno chiaramente, per Harris (1989), attribuire stati mentali ad altri, tramite simulazione, suppone l'attribuzione di suddetti stati mentali all'io sulla base dell'introspezione e, di conseguenza, i rilevanti concetti dello stato mentale. Stando alla tesi dell'introspezione, la simulazione può solamente essere

giustificata da un ragionamento analogico. Nella classica formulazione di questo argomento, “io sono internamente consapevole di una sequenza uniforme, il cui tassello iniziale si identifica con le modificazioni cui è soggetto il mio corpo, in un secondo stadio sopraggiungono i sentimenti ed infine si arriva al comportamento esterno. Nel caso degli esseri umani, io raggiingo l’evidenza dei sensi solo nel tassello iniziale e in quello finale, ma non in quello intermedio... sempre considerando che tale tassello sia della stessa natura... in questo modo posso esaminare altri esseri umani... secondo le stesse generalizzazioni” (J.S.Mill, *An examination of Sir William Hamilton’s Philosophy*, VI edizione, Londra, 1869).

Gordon difese una versione radicale della simulazione mentale, in base alla quale è necessaria una capacità di simulazione per raggiungere la vera capacità che consente di percepire gli oggetti come proiezioni mentali (e conseguentemente gli esseri umani come persone). Quantunque il compendio di informazioni psicologiche di cui disponiamo possa talvolta essere chiamato in causa nel quotidiano processo di «mind-reading», esso potrebbe anche risultare radicalmente inintelligibile per tutti quegli esseri incapaci di simulazione - forse proprio come la nozione di colore sfugge al daltonico. La mia rappresentazione di un altro in particolare è fondamentalmente una simulazione immagazzinata (forse come un insieme di modificazioni del mio io). Sebbene i dettagli della mia rappresentazione simulativa dell’altro siano passibili di giustificazione empirica, non v’è alcun modello inferenziale ulteriore (ad esempio, di origine analogica) tra la mia rappresentazione simulativa dell’altro e «le reali sembianze dell’altro».

Mentre nessun teorico della simulazione ha affermato che la nostra attività esplicativa e predittiva delle azioni altrui si basa sull’assunzione di un ruolo, Heal, con moderato successo, ha optato per un ibrido, simulazione-teoria, che vincola la simulazione primariamente a termini di matrice razionale, come credenze, desideri e azioni.

3. *L'argomento della parsimonia*

Secondo molti teorici, uno dei caratteri distintivi più rilevanti della psicologia del senso comune è il ruolo focale, essenziale che riveste il contenuto semantico degli stati postulati, soprattutto nel caso di atteggiamenti proposizionali indicativi quali possono essere le credenze, i desideri e le intenzioni. Si è affermato che la psicologia del senso comune comprende leggi o principi che vanno oltre il contenuto, legando, ad esempio, cosa *x* crede e cosa *x* desidera a cosa *x* sceglie di fare. Inoltre, tali legami sono comunemente considerati un riflesso delle relazioni semantiche tra i vari contenuti, relazioni che possono essere astrattamente rappresentate dai dettami della logica e del raziocinio come un *modus ponens* o un sillogismo pratico. Perciò la teoria della teoria prevede un inventario di leggi e principi causali che corrispondono a siffatti ruoli. Qui di seguito v'è un esempio sintomatico: cosa un individuo farà nel caso dovesse perdere la carta di credito, considerando le indesiderabili conseguenze che deriverebbero se non venisse compiuta tale azione:

Se *S* crede che se *p*, quindi *q* a meno che *S* non compia l'azione *x* e se *S* desidera il non-*q* e non crede che se *p* e *S* compiano l'azione *x*, allora a qualcosa fungerà *r* nel caso in cui *S* desideri il non-*r* più di quanto *S* desideri *p*, etc, etc; quindi, *ceteris paribus*, con ogni probabilità *S* compierà l'azione *x*.

Tuttavia, appare poco plausibile supporre un nostro eventuale richiamo a siffatti principi qualora volessimo anticipare il nostro comportamento in un futuro contingente o in situazioni contrafattuali. Per rispondere al quesito, «Cosa si farebbe se si dovesse perdere la carta di credito?», si potrebbe verosimilmente fare ricorso ad una strategia simulativa. Con ciò non si vuole affermare che tale strategia sia infallibile né negare il ricorso a generalizzazioni, ma queste ultime a volte sfociano in descrizioni non più teoretiche bensì comportamentali: ad esempio, «diventerò sempre fatalista in tali situazioni», dove «diventare fatalista» e «tali situazioni», sebbene interpretabili come descrizioni di matrice culturale, per la teoria

della teoria non sono descrizioni teoretiche: esse non descrivono il lavoro individuale del sistema comportamentale cognitivo e certamente non costituiscono leggi corrispondenti alle regole della logica e della ragione.

E proprio considerando le generalizzazioni causali come riflesso delle regole cui si adatta la nostra forma mentis, tale atteggiamento è reso ozioso proprio dalla teoria della simulazione. Indipendentemente dalle regole cui la nostra forma mentis si adatta, essa continua ad adattarsi a loro nell'ambito della simulazione, a meno che, ovviamente, non intervengano modificazioni per arginare le differenze. In altre parole, possiamo fare uso del nostro raziocinio come se fosse il prototipo del raziocinio di esseri che ragionano alla nostra stregua. Si deduce grazie a tale prospettiva che le ipotesi originate da un impianto speciale di leggi corrispondenti alle regole della logica e della ragione appaiono immotivate e prive di parsimonia.

4. Aree di ricerca empirica

Sono quattro le aree particolarmente rilevanti a tale proposito.

La *falsa credenza* ossia prendere in considerazione l'ignoranza o la falsa credenza altrui nel caso di predizione o spiegazione del loro stesso comportamento, processo che suppone, secondo la teoria della simulazione, modificazioni imaginative delle credenze individuali. Sono stati condotti numerosi test psicologici in modo da testare l'incapacità dei bambini di attribuzione di credenza. Ciò spiegherebbe anche la correlazione, in caso di autismo, dell'incapacità di attribuzione di credenza e l'incapacità di ricorrere spontaneamente ad un ruolo finto o anche all'interpretazione di un ruolo. Benché tutto ciò possa essere spiegato alla luce della teoria della teoria, la teoria della simulazione ne offre un'interpretazione alternativa (Wimmer e Perner, 1983; Gordon, 1986).

La *priorità dell'autoattribuzione* o dell'attribuzione all'altro. Si tratta di una seconda branca della ricerca sullo sviluppo in

base alla quale i bambini attribuiscono stati mentali a sé stessi prima di attribuirli ad altri. Varianti della teoria della simulazione che riconoscono i nostri stati mentali come tali e ricorrono ad inferenze analogiche di stati mentali altrui sembrano esigere una risposta affermativa a tale quesito; altre versioni invece della medesima teoria sembrano appoggiare una risposta negativa, avallata anche da alcuni test condotti. Ma il dibattito è ancora aperto (Gopnick e Wellman, 1992).

La *simulazione neuronale*: secondo la maggioranza delle versioni della teoria della simulazione, un rilevante quesito empirico, forse addirittura il quesito cruciale, è se i meccanismi neuronali e i processi di comprensione ed anticipazione del riscontro altrui nei confronti del mondo assomigliano in maniera significativa a quelli chiamati in causa per un riscontro in «prima persona» nei confronti del mondo. Si arriva ad un punto di convergenza: il cervello umano dispone di sistemi attivabili sia in modo endogeno – ad esempio, tramite sistemi propri di decisione-attrazione, emozione-formazione o nocipercezione - sia in modo esogeno in risposta alla visione diretta di facce e corpi umani. Tuttavia, i medesimi riscontri sono provocati anche quando la faccia altrui è sorpresa ad esprimere un'emozione corrispondente.

Inoltre, v'è una prova che danneggia la corteccia somatico-sensoriale che «legge» cosa accade nelle viscere di qualcuno, nel caso in cui si leda pesantemente l'abilità individuale di riconoscere l'emozione espressa sulla faccia di un altro. Perciò, la ricognizione di un'emozione espressa da un viso sembra fondarsi esclusivamente su tali riscontri viscerali esogeni o indotti dall'altro (Adolphs et al., 2000). Infatti, una grande quantità della nostra percezione del comportamento altrui si origina evidentemente grazie a tali duplici sistemi. Il processo neuronale alla base del nostro comportamento propositivo – ossia azioni come cercare una mela, afferrarla e portarla alla bocca per mangiarla - normalmente si verifica in seguito al processo decisionale. Eppure, una gran parte dei processi di organizzazione comportamentale accade in modo esogeno come se tali processi partecipassero ad un sistema di ricognizione percettivo. Recentemente sono stati scoperti

nella corteccia premotoria della scimmia macaco e, successivamente, anche nell'uomo, alcune cellule particolari, denominate neuroni «specchio», che possono attivarsi in due distinti contesti: sia quando l'animale esegue una particolare azione finalizzata nei confronti di un oggetto, sia quando osserva qualcun altro eseguire lo stesso tipo d'azione. Ad esempio, alcuni possono attivarsi quando si afferra un oggetto o quando si vede qualcuno compiere tale azione, sia con la mano sia con la bocca; altri invece sono più limitati e specifici, legati per esempio all'afferrare un oggetto solo con la mano o ad oggetti che si afferrano in un determinato modo.

È stato provato che, grazie a questi duplici sistemi, la vista di altri esseri (viventi) o di corpi simili al corpo umano immette nel nostro cervello non solo una rappresentazione visuale del loro comportamento, ma anche una sua replica interna, come, del resto, l'attività motoria o viscerale e riscontri di *nocipercezione* – ed eventualmente anche intenzioni minime – alla base del nostro comportamento. Rimane da verificare se e come tale simulazione *bottom-up* possa funzionare in combinazione con processi cognitivi, tra cui soprattutto i procedimenti immaginativi tante volte sottolineati da filosofi e psicologi e detti «*putting-in-place*»: ad esempio se e come i neuroni specchio possano «rivolgersi» ai processi cognitivi del processo decisionale altrui (Gallese e Goldman, 1998; Gordon, 2000).

L'impenetrabilità cognitiva. Per Stich e Nichols la simulazione è «impenetrabile da un punto di vista cognitivo», dato che essa opera indipendentemente da ogni conoscenza generale che un simulatore potrebbe avere sulla psicologia umana. Eppure, essi evidenziano che, quando i soggetti smarriscono certe informazioni psicologiche, possono talvolta operare previsioni sbagliate e perciò non si può più parlare di simulazione (Stich e Nichols, 1992). A causa di problemi metodologici ed interpretativi, come sottolineato da diversi filosofi e psicologi, la forza di persuasione di tale linea critica è oscura.

Alcuni filosofi pensano che la teoria della simulazione possa gettar luce su questioni topiche della filosofia della mente e del linguaggio concernenti l'intenzionalità, l'opacità refe-

renziale, contenuti ampi o limitati, la natura del processo causale della mente, i problemi Twin Earth e le peculiarità dell'autoconoscenza. Parecchi filosofi hanno applicato tale teoria all'estetica, all'etica ed alla filosofia delle scienze sociali. Eventuali successi o fallimenti di tali sforzi volti a scovare una risposta a problematiche filosofiche possono essere considerati test empirici della teoria, conferendo un senso quanto mai ampio al termine «empirico».

Bibliografia

Testi principali

- ADOLPHS, R. *et al.*, *A Role for Somatosensory Cortices in the Visual Recognition of Emotion as Revealed by Three-Dimensional Lesion Mapping*, in «Journal of Neuroscience», 20 (7), 2000, pp. 2683-2690.
- GALLESE, V., *The 'shared manifold' hypothesis: from mirror neurons to empathy*, in «Journal of Consciousness Studies», 8, 2001, pp. 33-50.
- GALLESE, V., GOLDMAN, A., *Mirror neurons and the simulation theory of mind-reading*, in «Trends in Cognitive Sciences», 2, 1998, pp. 493-501.
- GOLDMAN, A., *Interpretation Psychologized* in «Mind and Language» 4, 1989, pp. 161-185; ristampato in DAVIES, M., STONE T., *Folk Psychology: The Theory of Mind Debate*, Oxford, Blackwell Publishers, 1995.
- GOPNIK, A. and Wellman, H.M., 1992: "Why the Child's Theory of Mind Really Is a Theory," *Mind and Language* 7, 145-71.
- GORDON, R., *Folk Psychology as Simulation*, in «Mind and Language» 1, 1986, pp. 158-171; ristampato in DAVIES, M., STONE T. (a cura di), *Folk Psychology: The Theory of Mind Debate*. Oxford, Blackwell Publishers, 1995.
- , *Simulation Without Introspection or Inference From Me to You*, in «Mental Simulation: Evaluations and

Robert M. Gordon

- Applications», in M. DAVIES, T. Stone (a cura di), Oxford, Blackwell, 1995.
- GORDON, R., *Intentional Agents Like Myself*, in «Perspectives on Imitation: From Neuroscience to Social Science», vol. 2, a cura di S. Hurley e N. Chater, Cambridge, MA, MIT Press, 2004.
- HARRIS, P., *Children and Emotion*, Oxford, Blackwell Publishers, 1989.
- HEAL, J., *Replication and Functionalism*, in *Language, Mind, and Logic*, J. Butterfield (a cura di), Cambridge, Cambridge University Press, 1986; ristampato in DAVIES, M., STONE T. (a cura di), *Folk Psychology: The Theory of Mind Debate*, Oxford, Blackwell Publishers, 1995.
- HURLEY, S., *The Shared Circuits Hypothesis: A Unified Functional Architecture for Control, Imitation, and Simulation*, in «Perspectives on Imitation: From Neuroscience to Social Science», vol. 1, S. Hurley, N. Chater (a cura di), Cambridge, MA, MIT Press, 2004.
- STICH, S., NICHOLS, S., *Folk psychology: Simulation or tacit theory?*, in «Mind and Language» 7, 1992, pp. 35-71; reprinted in DAVIES, M., STONE T. (a cura di), *Folk Psychology: The Theory of Mind Debate*. Oxford, Blackwell Publishers, 1995.
- WIMMER, H., J. PERNER, *Beliefs About Beliefs: Representation and Constraining Function of Wrong Beliefs in Young Children's Understanding of Deception*, in «Cognition» 13, 1983, pp. 103-128.

Opere collettive:

- CARRUTHERS, P., SMITH, P., (a cura di), *Theories of Theories of Mind*. Cambridge, Cambridge University Press, 1996.
- DAVIES, M., STONE T. (a cura di), *Folk Psychology: The Theory of Mind Debate*, Oxford, Blackwell Publishers, 1995.
- , (a cura di), *Mental Simulation: Evaluations and Applications*. Oxford, Blackwell Publishers, 1995.

Folk Psychology

- DOKIC, J. and PROUST, J., (a cura di), 2002, *Simulation and Knowledge of Action*, Advances in Consciousness Research 45, John Benjamins, 2002.
- KOEGLER, H. STUEBER, K., (a cura di), *Empathy and Agency: the Problem of Understanding in the Social Sciences*, Westview Press, 2000.

Ulteriori contributi

- GOLDMAN, A., *The Psychology of Folk Psychology*, in «The Behavioral and Brain Sciences», 16, 1993, pp. 15-28.
- GORDON, R. M., J. BARKER, *Autism and the "theory of mind" debate*, in «Philosophical Psychopathology: A Book of Readings», G. GRAHAM, L. STEPHENS (a cura di), MIT Press, 1994, pp. 163-181.
- GORDON, R.M., 1995, *Sympathy, Simulation, and the Impartial Spectator*, in «Ethics», 105, pp. 727-742. Ristampato in *Mind and Morals: Essays on Ethics and Cognitive Science*, a cura di L. MAY, M. FRIEDMAN, A. CLARK, MIT Press, 1996.
- PEACOCKE, C., (a cura di), *Objectivity, Simulation, and the Unity of Consciousness*, Oxford, Oxford University Press, 1994.
- PERNER, J., *Understanding the Representational Mind*. Cambridge, MA, MIT Press, 1991.
- WELLMAN, H. M., *The Child's Theory of Mind*, Cambridge, MA, MIT Press., 1960

Appendice

COGNIZIONE E SIGNIFICATO *Lo sviluppo di forme categoriali nelle reti neurali dinamiche*

Enrica Vizzinisi

«...bisogna dunque essere in grado di comparare, di riflettere e di astrarre; queste tre operazioni logiche dell'intelletto, infatti, sono le condizioni essenziali e universali per la produzione di qualunque concetto in generale. Io vedo, ad esempio, un salice e un tiglio. Confrontando questi oggetti tra loro, innanzitutto, noto che essi sono diversi uno dall'altro riguardo al tronco, ai rami, alle foglie, ecc.; ma poi, riflettendo solo su ciò che essi hanno in comune fra loro: il tronco, i rami e le foglie stesse, e astraendo dalla loro grandezza, dalla loro figura, ecc., ottengo un concetto di albero» (I. Kant, *Critica della Ragion Pura*, Adelphi, Milano 1976).

Introduzione

Uno dei temi principali intorno al quale da tempo ruota l'attenzione di filosofi, linguisti e psicologi cognitivi è la capacità umana di «categorizzare». Come è agevole constatare, la facoltà di pensiero può, in effetti, considerarsi fondata sulla possibilità di una organizzazione categoriale della realtà. «Without concepts, mental life would be chaotic» (Smith e Medin, 1981: 1): l'idea centrale espressa da questa frase, frequentemente citata, è che i concetti sono talmente importanti per le funzioni cognitive che un pensiero senza concetti risulta difficilmente concepibile. Il mondo di cui abbiamo esperienza è costituito da una quantità di oggetti ed eventi così vasta che, se dovessimo individuarli tutti e nominarli singo-

larmente, saremmo sopraffatti dalla complessità dell'ambiente e della particolarità. La creazione di classi di oggetti o eventi simili consente, quindi, di gestire la complessità dell'esperienza fenomenologica, rendendo tutti gli organismi viventi (seppure in misura diversa e secondo modalità diverse) capaci di produrre risposte simili a classi di stimoli analoghi. Questa capacità di generalizzazione e di astrazione rappresenta un considerevole vantaggio competitivo, in quanto consente agli organismi viventi di interagire con il proprio ambiente con il minor dispendio possibile di risorse cognitive. Appare, quindi, evidente come una delle aree di indagine di maggiore interesse, su cui non potevano che concentrarsi gli sforzi dei ricercatori, sia quella relativa alla *concept formation*, intesa come processo sottostante la memorizzazione dei multiformi dati sensoriali, ma anche – in modo più o meno esplicito – di ogni insieme di entità concrete o astratte che devono essere elaborate in modo analogo ai dati sensoriali.

A partire dagli anni '70, lo studio dei concetti e delle categorie ha suscitato un rinnovato interesse da parte degli psicologi, che hanno cominciato a dedicarsi alla formulazione di una serie di teorie sulla concettualizzazione, complementari o alternative tra loro. Il problema della categorizzazione, così come viene posto dagli attuali scienziati cognitivisti, non può, in effetti, che richiamare alla mente una serie di problematiche tradizionalmente discusse nell'ambito della filosofia della conoscenza e del linguaggio. Tra queste problematiche le principali possono essere riassunte come segue (Harnard, 1987):

- il problema dell'induzione: come si arriva a formulare delle generalizzazioni corrette e predittive, a partire da un insieme finito di esempi particolari?
- il problema dell'attribuzione di un nome agli oggetti (nominalismo): che relazione intercorre tra le parole dei linguaggi naturali e gli oggetti che esse designano?
- il problema del significato olistico: il significato delle singole parole è indipendente dal resto del vocabolario, oppure ogni parola è in qualche modo vincolata al resto del vocabolario,

in modo tale che, cambiando il significato associato ad essa, si renda necessario revisionare l'intero bagaglio lessicale?

– il problema della conoscenza–gnoseologica e della conoscenza-linguistica: qual è la differenza tra ciò che sappiamo (o potremmo sapere) direttamente dall'esperienza sensoriale e ciò che sappiamo (o potremmo sapere) dalle descrizioni verbali?

– il problema dei concetti elementari: quali sono le unità sulla base delle quali costruiamo la nostra percezione degli oggetti e delle forme? Queste unità si possono considerare come dei primitivi percettivi indivisibili?

– il problema dei simboli atomici: in che modo le parole individuali che usiamo nelle nostre definizioni e descrizioni derivano il loro significato? Come è possibile, in altre parole, evitare il regresso all'infinito della definizione di termini per mezzo di termini?

– il problema degli universali: che differenza esiste tra un'istanza particolare di un oggetto e le sue caratteristiche (universali)? In che modo si possono rappresentare le proprietà concrete vs. le proprietà astratte?

Nessun'altra funzione cognitiva appare, in questo modo, più adatta della categorizzazione per spiegare i principi di base che caratterizzano la cognizione in senso generale. Alla base della capacità umana di creare categorie si troverebbe, in effetti, una precisa struttura mentale (la struttura concettuale) in cui «l'informazione linguistica, l'informazione sensoriale, e l'informazione motoria divengono compatibili» (Jackendoff, 1983; Jackendoff, 1987), e in cui la generazione semantica diviene il prodotto di un'azione congiunta di diverse funzioni mentali: percezione, linguaggio, rappresentazione della conoscenza, memoria, ecc..

La definizione di modelli formali della categorizzazione rappresenta così un punto di partenza fondamentale per il progetto più ambizioso di giungere ad una completa modellizzazione e comprensione della mente. A tale scopo, nell'ambito delle scienze cognitive e della filosofia della mente sono stati messi appunto, nel corso degli anni, diversi mo-

delli per la formazione di concetti, raggruppabili nei due seguenti tipi:

Modelli insiemistici della concettualizzazione. In accordo a questi modelli i concetti vengono definiti sulla base della loro appartenenza ad una classe di oggetti caratterizzata da un insieme di condizioni necessarie e sufficienti. Il significato di un'espressione complessa è dato in funzione del significato delle sue parti (proprietà) e del modo in cui tali parti si combinano sintatticamente tra di loro (relazioni) [principio della composizionalità (PC) di Frege]⁵³.

Modelli concettuali basati su prototipi, secondo i quali la rappresentazione cognitiva associata a molte parole di uso comune (parole come «mobile», «frutto», «vestito», ecc.) richiama alla mente un «buon esempio» della categoria in questione. L'appartenenza di un'occorrenza particolare ad una categoria viene, quindi, stabilito sulla base della distanza metrica dal dominio concettuale di appartenenza, cioè, da una funzione binaria che assegna ad ogni coppia di elementi un valore numerico che indica la misura in cui tali elementi differiscono tra di loro e da una funzione che assegna ad ogni elemento un numero nell'intervallo reale $[0,1]$, numero che indica il «grado» in cui tale elemento appartiene ad un concetto⁵⁴

In effetti, però, nessuno di questi due tipi di modellizzazione è riuscito a definire a pieno la generatività semantica e concettuale che caratterizza la mente umana. La capacità di categorizzare appare, infatti, strettamente vincolata ad una componente attiva della cognizione che continuamente cerca di adattare e di riorganizzare la struttura concettuale nel tentativo di rendere massima la stabilità della struttura globale (sistemi con regole di preferenza) (Jackendoff, 1987). Tale componente funziona sulla base di alcuni principi innati dell'organizzazione percettiva, maturati nell'ambito della Scuola della Gestalt, quali, ad esempio, i fattori di prossimità,

⁵³ Cfr. Braisby, 1998; Katz, J. & Fodor, 1967; Frege, G. (1892/1980)

⁵⁴ Cfr. Kamp. H. Partee, B., 1995; Rosch, E., 1962

vicinanza, chiusura, ecc. In questo quadro, il problema di fondo che viene a profilarsi all'orizzonte è dunque quello di giungere ad individuare la forma «giusta» delle leggi dinamiche che determinano l'articolazione delle strutture cognitive. Non si tratta più soltanto di definire distanze metriche, o postulati di significato, come nella tradizione compositiva, si tratta di considerare, bensì, in profondità gli aspetti propri del contesto semantico, nonché del contesto dinamico al cui interno giungono ad articolarsi le procedure della classificazione e della categorizzazione che caratterizzano più propriamente le attività cognitive. Occorre definire, in altre parole, tipi di algoritmi capaci di generare oggetti dotati di un tipo particolare di organizzazione: l'organizzazione intensionale. Oggetti capaci di autorganizzarsi e di articolarsi in modo autonomo, anche sulla base della conoscenza ambientale-specifica acquisita per mezzo degli organi sensoriali. È così che, in contrapposizione al modello della mente computazionale tradizionale (HIP, Human Information Processing), in cui l'informazione rappresenta la componente «passiva», viene sempre più emergendo un ruolo «attivo» dell'informazione. L'interazione mente-mondo da origine a dei processi di apprendimento, che spesso consistono nella costruzione di nuove «strutture mentali», cioè, di nuovi «legami» (associazioni) che potrebbero anche corrispondere a delle specifiche strutture neurali.

Per far fronte a queste considerazioni si è sviluppata di recente una linea di ricerca che centra la propria attenzione soprattutto sul modo in cui i dati sensoriali vengono codificati internamente, anche grazie all'interazione di agenti artificiali con il mondo in cui sono inseriti. A partire dal concetto di *symbol grounding problem* (Harnard, 1990), uno dei principali problemi incontrati dalla metodologia simulativa in psicologia, alcuni ricercatori hanno sviluppato diversi modelli di reti neurali per la categorizzazione percettiva⁵⁵ Nell'ambito di questi nuovi modelli, la formazione di concetti viene a deli-

⁵⁵ Cfr. Greco A., Cangelosi A., Harnard S. (1998), Murre J.M.J., Phaf, R.H., & Wolters, G. (1992), Schyns P.G. (1991)

nearsi come un fenomeno emergente in cui la semantica degli elementi è assegnata in seguito all'interazione fra un agente e il suo ambiente (semantica situazionale)⁵⁶. Resta quindi valida l'ipotesi che esista una corrispondenza fra oggetti del mondo e una loro codifica interna, ma si vuole approfondire come viene assegnato il significato alle codifiche interne in modo da poter dire che queste rappresentano aspetti del mondo (*symbol grounding problem*). Ciò permetterebbe di giustificare e modellizzare la flessibilità che si fa nell'uso dei concetti, flessibilità derivata dall'influenza del contesto sulla loro semantica. Nell'ambito di questo nuovo approccio al problema della *concept formation*, sono stati assunti dei principi di base. Innanzitutto è stato assunto che il mondo non contiene informazioni ma dati, che è l'agente stesso a trasformare in informazioni. L'informazione viene, in questo modo, a costituire il risultato di un processo di interazione con il mondo e quindi è relativa e soggettiva. L'informazione è tale solo se permette all'agente di svolgere più efficacemente le attività, volte al raggiungimento di qualche obiettivo, in cui esso è, o è stato coinvolto. Le modalità di estrazione dei dati hanno poi un impatto profondo su tutti i processi cognitivi, anche di alto livello, che non possono prescindere dal modo in cui le informazioni stesse sono state formate. Il processo di apprendimento è pertanto continuo, l'agente è in costante interazione con il mondo e aggiorna continuamente le proprie rappresentazioni interne degli aspetti del suo ambiente rilevanti rispetto al contesto e agli obiettivi correnti.

Una prima nozione di agente situato, ossia di sistema in grado di adattarsi con successo all'ambiente, esibendo comportamenti essenzialmente reattivi ed evitando l'uso di rappresentazioni simboliche e di complessi processi di inferenza, si ha in un articolo di Rodney Brooks (Brooks, 1991). Condi-

⁵⁶ Cfr. Barwise, J. & Perry, J. (1983), *Situations and Attitudes*, Cambridge Mass.

zione necessaria di tale successo adattivo è costituito secondo Brooks dalla fitta e continua rete di interazioni tra l'agente ed il proprio ambiente, interazioni che consentono allo stesso agente di evitare la costruzione di modelli del mondo in un formato simbolico/rappresentazionale per «affidarsi», per così dire, alle risorse sensoriali offerte dall'ambiente stesso. Tali risorse, nel caso dei sistemi progettati da Brooks, si configurano come un insieme di stimoli sensoriali locali in grado di attivare risposte comportamentali predefinite. Il punto di partenza di Brooks è la critica rivolta alle architetture cognitive tradizionali, in quanto gravate di un carico computazionale eccessivo. Tale carico è costituito dalle rappresentazioni e dai processi di inferenza, che impedirebbero la produzione di risposte adattivamente efficaci ed in tempo reale alle situazioni problematiche che vengono a presentarsi nell'ambiente esterno, fornendo così un modello alquanto implausibile delle attività cognitive di base di un sistema. Queste attività, diversamente da quanto prospettato dalle ricerche dell'IA classica, non sarebbero rappresentate dai processi di ragionamento logico e di pianificazione, ma dalla capacità da parte di un sistema di interagire strettamente con l'ambiente, proprio in vista della riduzione del carico computazionale individuale, selezionando «invarianti sensoriali» nell'ambiente stesso. La riduzione di tale carico, nei modelli di agente situato, è realizzata riducendo il percorso di calcolo complessivo che porta dalla sensazione al comportamento, eliminando i processi di codifica dell'informazione in rappresentazioni simboliche e di inferenza logica, in favore di un più stretto legame tra percezione e azione. Gli agenti in questione sarebbero quindi guidati dall'ambiente stesso nella produzione dei comportamenti attraverso la rilevazione di stati parziali ed indicatori di configurazioni sensoriali in grado di attivare risposte adeguate. L'agente sarebbe situato nel proprio ambiente in quanto coinvolto in cicli continui di percezione-azione-nuova percezione che consentono un adattamento efficace ed allo stesso tempo consentono una rinnovata concezione dell'ambiente stesso inteso non più come semplice «deposito» di informazioni o come scenario passivo delle azioni degli agenti cogni-

tivi, ma come fattore paritario all'interno del ciclo computazionale complessivo che porta alla produzione dei comportamenti.

1. *Alcuni esempi di reti neurali modulari
per l'acquisizione di concetti*

Tra i modelli formali della cognizione elaborati negli ultimi decenni, quelli che più sembrano essere di supporto alla implementazione di agenti situati sono le reti neurali modulari. Tali reti sono caratterizzate da più moduli, ognuno dei quali porta a termine un compito specifico, elaborando l'informazione in maniera isolata, interfacciandosi, cioè, con gli altri moduli solo a livello di input ed output. Mentre i pesi che connettono i vari moduli sono modificabili, gli stessi moduli, una volta che hanno portato a termine i propri cicli di apprendimento, sono fissati strutturalmente e funzionalmente. Nella maggioranza di questi modelli le diverse funzioni da svolgere consistono sostanzialmente nella generazione di prototipi, a partire da una serie di pattern (vettori input), che vengono somministrati alla rete e nell'attribuzione di un nome a tali prototipi. Queste funzioni vengono solitamente portate a termine per mezzo di reti neurali diverse, caratterizzate, cioè, da architetture e modalità di apprendimento specifiche. In queste reti, solitamente, il problema della concept formation, viene affrontato partendo da un sub-problema, quello della percezione categoriale. Attraverso questa funzione cognitiva è possibile estrapolare dalla rappresentazione iconica (tracce sensoriali) di un oggetto le caratteristiche invarianti, quelle caratteristiche, cioè, che lo rendono uguale agli oggetti appartenenti alla stessa classe e differente da tutti gli altri. Fa seguito a questa prima fase, il raggruppamento delle stesse caratteristiche in «chunks» o categorie. La seconda fase costituisce propriamente la percezione categoriale, volta a massimizzare la differenza fra i confini percepiti e a minimizzare quella al loro interno. In effetti, le differenze fisiche di ugual misura, esistenti tra gli stimoli provenienti dall'esterno tramite i nostri recettori sensoriali, sono percepite come più picco-

le nell'ambito delle stesse categorie e come più grandi tra categorie diverse.

La percezione categoriale è una funzione cognitiva dovuta in parte all'esperienza ed in parte acquisita. Essa riguarda diverse capacità cognitive: dal riconoscimento dei colori, alla percezione del parlato. Queste stesse capacità possono essere facilmente riprodotte da particolari reti neurali addestrate a rilevare le caratteristiche invarianti in pattern di input sensoriali, ad ordinare tali caratteristiche in modo da generare delle vere e proprie categorie ed infine ad attribuire a queste ultime un nome. Tali reti si basano su alcuni principi di funzionamento propri del sistema nervoso, quali, ad esempio, la modularità ed un'organizzazione con connessioni inibitorie ed eccitatorie. Imporre una struttura modulare a delle reti neurali significa dotare queste stesse reti di un'architettura iniziale più elaborata, capace di produrre, attraverso cicli di addestramento, una struttura che presenta maggiore stabilità delle rappresentazioni apprese ed una ridotta interferenza tra apprendimenti successivi. È questo ciò che si verifica in alcune delle reti neurali recentemente sviluppate⁵⁷.

In particolare, il modello di Schyns è caratterizzato da tre diversi tipi di reti neurali. La formazione delle rappresentazioni concettuali avviene per mezzo di una mappa auto-associativa di Kohonen. L'associazione alla categoria di un nome, non essendo possibile senza una forma di apprendimento supervisionato, si realizza, al contrario, tramite un auto-associatore con regola di apprendimento basato sull'errore, del tipo di Widrow-Hoff (Hinton, 1987). Una volta che l'apprendimento ha avuto luogo, uno schema di richiamo BSB (Brain State in a Box)⁵⁸ classifica un'istanza di una cate-

⁵⁷ Cfr. Schyns, (1991); Murre J.M.J., Phaf, R.H., & Wolters, G. (1992); Greco A., Cangelosi A, Harnard S. (1998)

⁵⁸ Cfr. Anderson, J.A., Silverstein, J.W., Ritz, S.A. e Jones, R.S. (1977), "Distinctive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model", in *Psychological Review*, vol. 84, pp.413-451

goria particolare, secondo l'interpretazione concettuale adeguata, e le associa un nome.

Un po' più complesso risulta il modello CALM⁵⁹, basato su moduli ad apprendimento non supervisionato di tipo competitivo⁶⁰. Un modulo per la categorizzazione e l'apprendimento è costituito da quattro gruppi funzionali di nodi mutuamente esclusivi. L'informazione entra in un modulo attraverso i nodi per la rappresentazione, gli R-Nodi. Gli altri tre nodi presenti in CALM (i V-nodi, gli A-nodi ed, infine, gli E-nodi) implementano, successivamente, il meccanismo competitivo che fa sì che alla fine ci sia sempre un R-nodo vincitore per ciascun pattern input.

Infine, un altro modello particolarmente rilevante è quello di Greco A., Cangelosi A., Harnard S. (1998), basato su apprendimento per back-propagation. A differenza dei modelli neurali fin qui trattati, in quest'ultimo modello la modularità è di tipo essenzialmente funzionale. In altre parole, a differenza del modello di Schyns o del modello CALM, in cui le diverse funzioni venivano supportate da architetture differenziate e autonome, in questo modello i diversi compiti cognitivi vengono portati a termine da una stessa rete che esegue in successione tre diversi cicli di apprendimento: ricerca del prototipo; attribuzione del nome a livello dello stimolo in entrata; attribuzione del nome a livello più elevato.

2. I limiti del connessionismo e l'ipotesi dinamica

Nonostante i primi successi ottenuti dall'applicazione di queste reti neurali, ancora molta è, tuttavia, la strada che ci separa dal poter considerare tali sistemi connessionistici dei modelli adeguati per l'acquisizione di concetti e, in termini

⁵⁹ Cfr. Murre J.M.J., Phaf, R.H., & Wolters, G. (1992).

⁶⁰ Tali modelli si sono sviluppati agli inizi degli anni '70 grazie al contributo di Christoph von der Malsburg (von der Malsburg, 1973) e Stephen Grossberg (Grossberg, 1976) e si sono poi evoluti grazie a ricercatori quali, ad esempio, Shun-Ichi Amari (Amari, 1978) e Teuvo Kohonen (Kohonen, 1984).

più generali, per la generazione di strutture semantiche. Queste reti presentano, infatti, caratteristiche molto interessanti: 1) dall'attività congiunta dei moduli e sulla base di un controllo distribuito sulle reti stesse, emergono dei pattern dinamici di attivazione; 2) i processi di apprendimento sono continui ed auto-organizzanti, in quanto collegati ai processi percettivi. Ma queste stesse caratteristiche non sono però sufficienti a spiegare la complessità dei processi cognitivi implicati nella formazione di strutture concettuali nell'uomo. Quello che appare sfuggire a tali sistemi neurali è, infatti, la creatività tipica del pensiero umano, la spontaneità avrebbe detto Kant, delle procedure proprie della mente. Quella stessa creatività e metamorfosi che porta l'uomo a creare dei concetti ambigui, delle sottili sfumature di significato o quei «giochi» di parole che le metafore e le metonimie esprimono a livello dei linguaggi naturali. Al fine di riprodurre artificialmente tali capacità sarebbe, quindi, necessario ricorrere a sistemi capaci di gestire l'associazionismo concettuale e di operare in un interazionismo neurale in cui ogni operazione venga svolta tenendo conto dell'intera vita mentale dell'organismo. Bisognerebbe, in altre parole, giungere a definire meglio come la conoscenza delle operazioni di classificazione acquisite in questi ultimi anni, a livello di simulazione tramite reti neurali, possa connettersi agli aspetti tipicamente dinamici che sembrano caratterizzare, così come tutti i sistemi complessi, anche la cognizione.

In effetti, anche se l'applicazione delle reti neurali ha fatto ottenere dei buoni risultati sul piano della simulazione cognitiva, è, comunque, bene tenere in considerazione il fatto che, l'elevata non-linearità dei sistemi complessi, non consente per il momento di predire il comportamento dell'intera rete, a partire dalla conoscenza delle regole di interazione locali. Considerare i fenomeni cognitivi come derivanti dall'attività di sistemi dinamici significa, infatti, valutare tutte le variazioni quantitative riguardanti gli stati dei subsistemi che costituiscono il sistema cognitivo nella sua interezza. Per passare da un livello di rappresentazione basilare ad un livello più elevato occorre, così, inserire le reti modulari, del tipo qui de-

scritte, in un contesto più ampio. In un contesto, in particolare, dove la cognizione venga intesa non più come un dispositivo astratto che trasforma input simbolici in output simbolici, ma piuttosto come un tutto integrato, dove le funzioni cognitive elevate risultano essere il prodotto dell'attività congiunta e mutevole nel tempo di tre sistemi complessi: il sistema nervoso, il corpo e l'ambiente esterno. Un sistema cognitivo di questo tipo non interagisce, quindi, con il corpo e l'ambiente esterno attraverso input ed output simbolici occasionali e statici: questa interazione appare, al contrario, come un vero e proprio accoppiamento strutturale (Maturana e Varela, 1980), tale che tutti i processi coinvolti vengano continuamente e reciprocamente influenzati in direzione di una simultanea e mutua coevoluzione.

The world thus perceived at the visual level is constituted not by objects or static forms, but by processes appearing imbued with meaning. As Kanizsa stated, at the visual level the line per se does not exist: only the line which enters, goes behind, divides, etc.: a line evolving according to a precise holistic context, in comparison with which function and meaning are indissolubly interlinked. The static line is in actual fact the result of a dynamic compensation of forces. Just as the meaning of words is connected with a universe of highly-dynamic functions and functional processes which operate syntheses, cancellations, integrations, etc. (a universe which can only be described in terms of symbolic dynamics), in the same way, at the level of vision, I must continuously unravel and construct schemata; must assimilate and make myself available for selection by the co-ordinated information penetrating from external reality. Lastly, I must interrelate all this with the internal selection mechanisms through a precise «journey» into the regions of intensionality (Carsetti, 2004; p. 325)

Alla luce di queste considerazioni, la strada che ci separa dalla realizzazione nel futuro di modelli della cognizione sempre più adeguati appare molto lunga e la costruzione di simulazioni adatte ai processi cognitivi un compito difficile. In ef-

fetti, la realtà che si intende modellizzare (la mente) presenta al contempo molteplici aspetti: è adattiva, evolvente, dissipativa, auto-referenziale, ecc. Possiede una struttura linguistica e si articola secondo un preciso contesto semantico.

Ad ogni modo, appare evidente che i primi e unici due aspetti che caratterizzano questa complessa realtà sono la dinamicità e l'auto-organizzazione. È proprio a partire da questi due aspetti che di recente diversi ricercatori hanno delineato un nuovo approccio allo studio della mente: l'ipotesi dinamica⁶¹.

Per i sostenitori di questa ipotesi il cervello è un ampio e complesso sistema fisico caratterizzato da un'azione dinamica che si svolge in una continuità temporale e che consiste nell'eccitazione e nell'inibizione delle cellule neurali. Il comportamento di un numero elevato di microscopici elementi neurali determina l'emergenza di bacini di attrazione discreti. Questi macro stati dinamici costituiscono il substrato per gli stati mentali e comportamentali. In particolare, i bacini di attrazione rappresentano soltanto una tra le entità tipiche dei Sistemi Dinamici, le quali includono l'assegnazione di parametri, gli stati del sistema, gli attrattori, le traiettorie, o perfino aspetti delle strutture di biforcazione. In queste entità gli studiosi di orientamento dinamico ritrovano le rappresentazioni interne, quelle stesse rappresentazioni, cioè, che gli studiosi di orientamento computazionale, identificavano nelle configurazioni statiche di occorrenze simboliche (*symbol tokens*). Un'ovvia caratteristica che distingue i modelli dinamici nelle scienze cognitive dai modelli computazionali standard è, infatti, che le loro variabili sono numeriche.

⁶¹ Tale approccio si ritrova in delle reti neurali dinamiche: (Beurle 1956; Greene 1962; Rashevsky 1960), nella psicologia della Gestalt (Köhler 1951), nella psicologia ecologica di stampo gibsoniano (Carello et al. 1984), nella modellizzazione basata su EEG (Basar 1989; Nunez 1995), ed, infine, nella recentissima teoria dei sistemi dinamici (Freeman 1975; Freeman 1995; Freeman 1999; Haken 1983; Haken 1991; Kelso 1995; Kugler 1987; van Gelder and Port 1995).

Come è agevole constatare, un sistema di questo tipo è, infatti, costituito da un insieme di parti correlate, percepite, però, come un tutto unitario. Lo stato del sistema ad ogni istante t è definito dall'insieme dei valori associati alle singole parti in quello stesso istante t . Nei sistemi dinamici, alcuni o tutti questi valori possono cambiare nel tempo: «Il termine dinamico si riferisce ai fenomeni che producono modelli di cambiamento temporale (...) il termine è quasi sinonimo di evoluzione temporale o schema di cambiamento»⁶². Ad ogni dato istante, lo stato del sistema costituisce un punto nello spazio metrico, che al passare del tempo, spostandosi, da origine ad una traiettoria. Le equazioni differenziali rappresentano, così, un modo matematico per studiare i sistemi dinamici. Tali equazioni descrivono, infatti le relazioni di cambiamento tra le variabili e la loro soluzione è data da una descrizione esplicita del modo in cui le variabili del sistema variano in funzione della variabile tempo (t). In molti sistemi complessi presenti in natura è impossibile arrivare ad una soluzione dell'equazione differenziale che li sottende. Quello che oggi, anche grazie all'ausilio tecnologico offerto dalle simulazioni computerizzate, si può ottenere sono, però, delle previsioni verosimili sul comportamento futuro del sistema. Ad esempio, il sistema tende ad un equilibrio stabile, oppure, presenta dei comportamenti periodici? Le traiettorie tendono a convergere oppure a divergere? Ci sono delle parti dello spazio in cui il sistema non transita mai? Ci sono degli attrattori verso i quali le traiettorie convergono?

Lo studio dei sistemi dinamici inizialmente applicato ai fenomeni di tipo fisico, sta sempre più prendendo piede anche nell'ambito delle scienze cognitive ed in particolare nell'ambito delle tradizionali problematiche associate alla generazione del significato. (Hardy, 1997; Hardy, 1998). Come già detto nel corso di questo articolo, la generatività semantica appare legata all'attività di una mente intesa come sistema complesso, in cui tutte le parti contribuiscono alla definizione

⁶² Cfr. Luenberger, D. G. (1979), p.1.

di una coerenza globale e alla determinazione di un'individualità dello stesso sistema. Non solo, contribuisce alla generazione del significato anche l'accoppiamento strutturale con l'ambiente e con gli altri organismi (Maturana e Varela, 1980). Una teoria completa della mente dovrebbe, quindi, tenere in considerazione, non solo l'interazione sistematica mente-psiche-corpo, ma anche quella mente-nel-mondo.

È questo quello che si accenna a realizzarsi, grazie all'opera di tutti coloro impegnati nella definizione di una teoria dei campi semantici (cfr. Hardy, 1998), nel campo della Sinergica (Haken, 1983), delle reti neurali, della psicologia ecologica, della robotica situata e della vita artificiale. L'azione congiunta di studi apparentemente distanti tra di loro sta, in effetti, gradualmente definendo un modello della mente basato su un lattice di numerose costellazioni semantiche (rappresentate da reti neurali di tipo dinamico), in grado di correlare ed associare un'ampia gamma di differenti elementi e processi (quali, ad esempio, concetti, sensazioni, azioni, parole, ricordi, ecc.) in un «meaningful whole».

Bibliografia

- ABRAHAM, F.D., ABRAHAM, R.H., SHAW, C.D., *Basic Principles of Dynamical Systems* in R. L. LEVINE, H.E. FITZGERALD (a cura di), *Analysis of Dynamic Psychological Systems, Volume 1: Basic Approaches to General Systems, Dynamic Systems, and Cybernetics*, Plenum Press, New York, 1992
- AMARI, S., TAKEUCHI, A.S., *Mathematical Theory on Formation of Category Detecting Nerve Cells*, *Biological Cybernetics*, Vol. 29, 1978, pp. 127-136
- ATLAN, H., *Self-organizing Networks. Weak Strong and Intensional*, in «La Nuova Critica», 1992, 19-20.
- BARWISE, J., PERRY, J., *Situations and Attitudes*, Cambridge Mass, 1983

- BASAR, E., *Brain natural frequencies are causal factors for resonances and induced rhythms*, in *Brain Dynamics*, E. Basar and T.E. Bulloch (a cura di), Springer-Verlag, Berlin 1989, pp. 425-57.
- BEURLE, R. L., *Properties of a mass of cells capable of regenerating pulses*, *Phil. Trans. Roy. Soc. London* 1956, pp. 55-94.
- BRAISBY, N., *Compositionality and the Modelling of Complex Concepts*, in «*Minds and Machines*», Vol. 8 n. 4, 1998, pp. 479-508
- BROOKS, R.A., *Intelligence without representation*, in «*Artificial Intelligence*», 1991, 47, pp. 139-159
- CARELLO, C., TURVEY, M. T., KUGLER, P. N., SHAW, R. E., *Inadequacies of the computer metaphor*, in «*Handbook of Cognitive Neuroscience*», M.S. Gazzaniga (a cura di), Plenum Press, New York, 1984, pp. 229-48.
- CARPENTER, G. A., GROSSBERG S., *ART 2: self-organization of stable category recognition codes for analog input patterns*, *Applied Optics*, 26, 1987, pp. 4919-4930.
- CARSETTI A., *Meaning and Complexity: a Non-Standard Approach*, in «*La Nuova Critica*» 19/20, 1992, pp. 109-126
– *Semantics Structures and Cognitive Functions*, in «*La Nuova Critica*» 30, 1997, pp. 99-147
– (a cura di), *Functional Models of Cognition. Self-Organising Dynamics and Semantic Structures in Cognitive Systems*, Kluwer Academic Publishers, Dordrecht, 1999
– (a cura di), *Seeing, Thinking and Knowing. Meaning and Self-Organisation in Visual Cognition and Thought*, Kluwer Academic Publishers, Dordrecht, 2004
- CHURCHLAND, P.S., SEJNOWSKI, T.S., *The Computational Brain*, MIT Press, Cambridge, 1995
- FREEMAN, W. J., *Mass Action in the Nervous System*, Academic Press, New York, 1975
–, *Societies of Brains. A Study in the Neuroscience of Love and Hate*, Lawrence Erlbaum, New York, 1995
–, *Consciousness, intentionality, and causality*, in «*Reclaiming Cognition*», ripubblicato in «*Consciousness Studies*», Vol. 6, Nos. 11-12, 1999.

- FREGE, G., (1892), *On sense and meaning*, in P. GEACH, M. BLACK (a cura di), *Translations from the Philosophical Writings of Gottlob Frege*, N.J., Barnes & Noble Books, Towota, 1980, pp. 56-78,
- GRECO A., CANGELOSI A, HARNARD S., *A connectionist model for categorical perception and symbol grounding*, in *Proceedings ICANN98 – International Conference on Artificial Neural Networks*, 1998
- GREENE, P. H., *On looking for neural networks and “cell assemblies” that underlie behaviour*, in «I. Mathematical model. II. Neural realization of a mathematical model *Bull. Math. Biophys.*», 1962, 24, 247-75, 395-411
- GROSSBERG, S., *Adaptive Pattern Classification and Universal Recoding, 1: Parallel Development and Coding of Neural Feature Detectors*, in «*Biological Cybernetics*», Vol. 23, 1976, pp.121-134
- HAKEN, H., *Synergetics. An Introduction*, Berlin-New York, Springer, 1983
- , H., *Synergetic Computers and Cognition*, Springer-Verlag, Berlin, 1991
- HARNARD, S. (a cura di), *Categorical Perception*, Cambridge, Cambridge University Press, 1987
- , *The Symbol Grounding Problem*, in «*Physica*» D, 42, 1990, pp. 335-346
- HARDY, C., *Semantic fields and meaning. World Futures*, 48, 161- 70, Newark, Gordon & Breach, 1997
- , *Networks of meaning: A bridge between mind and matter*, Westport, CT, Praeger, 1998
- JACKENDOFF, R., *Semantics and Cognition*, MIT Press, 1983
- , *Consciousness and the Computational Mind*, Cambridge, The MIT Press, 1987
- KAMP. H. PARTEE, B., *Prototype Theory and Compositionality*, in «*Cognition*» vol. 57 n.2, , pp. 129-191, 1995
- KATZ, J., FODOR, J., *The Structure of a Semantic Theory*, in Block, N. (a cura di), «*Reading in the Psychology of Language*», Englewood Cliffts, Prentice-Hall, 1967, pp. 398-431

- KELSO, J. A. S., *Dynamic Patterns: The Self-Organization of Brain and Behavior*, MIT Press, Cambridge, 1995
- KÖHLER, W., *Relational determination in perception*, in «Cerebral Mechanisms in Behavior: The Hixon Symposium», L.A. Jeffress (a cura di), Wiley, New York, 1951, pp. 200-43.
- KOHONEN, T., *Self-organized formation of topologically correct feature maps*, in «Biological Cybernetics», 43, 1982, pp. 59 - 69.
- , *Self-organization and associative memory*, Springer Verlag, 1984
- KUGLER, P. N. a. T., M.T., *Information, Natural Law, and the Self-assembly of Rhythmic Movement*, Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1987
- LAKOFF, G., *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*, University of Chicago, Chicago, 1987
- LUENBERGER, D. G., *Introduction to Dynamic Systems: Theory, Models, and Applications*, John Wiley & Sons, New York, 1979
- MATURANA, H., VARELA, F., *Autopoiesis and Cognition*, Dordrecht-Boston, 1980
- MCCULLOCH, W., PITTS, E., *A Logical Calculus of the Ideas Immanent in Neural Nets*, in «Bull. Math. Biophys.», 5, 115, 1943
- MURPHY, G.L., ROSS, B.H., *Prediction from Uncertain categorizations*, in «Cognitive Psychology» vol 27, n.2, 1994, pp. 148-193
- MURRE J.M.J., PHAF, R.H., WOLTERS, G., *CALM: Categorizing and Learning Module*, in «Neural Networks», Vol 5, 1992, pp. 55-82
- NUNEZ, P. L., *Towards a physics of neocortex*, in *Neocortical Dynamics and Human EEG Rhythms*, P.L. Nunez (a cura di), Oxford University Press, New York, 1995, pp. 68-132.
- PARISI, D., *Mente. I Nuovi Modelli di Vita Artificiale*, Il Mulino, Bologna 1999

- PATEL, M.L., SCHNEPF, U., *Concept Formation as Emergent Phenomena*, in «European Conference on Artificial Life», Cambridge, MIT Press, 1992
- PUTNAM, H., *Renewing Philosophy*, Harvard College, 1992
- RASHEVSKY, N., *Mathematical Biophysics: Physico-Mathematical Foundations of Biology*, Vols. I & II, Dover, New York, 1960
- ROSCH, E., *Cognitive Representation of Semantic Categories*, in «Journal of Experimental Psychology: General», 104, 1962, pp. 2-233
- RUMELHART, D.E., MCCLELLAND, J.L. (a cura di), *Parallel Distributed Processing. Explorations in the Microstructures of Cognition*, Volume 1, Foundations, Cambridge Mass., Mit Press, 1986
- SCHYNS P.G., *A modular neural network for concept acquisition*, in «Cognitive Science», 15, 1991, pp. 461-508
- SMITH, E.E., MEDIN, D.L., *Categories and Concepts*, Cambridge MA, Harvard University Press, 1981
- VON DER MALSBERG, C., *Self-Organization of Orientation Sensitive Cells in the Striate Cortex*, in «Kybernetik», vol.14, 1973, pp. 85-100
- VAN GELDER, T., PORT, R. F. (a cura di), *Mind as Motion: Explorations in the Dynamics of Cognition*, MIT Press, Cambridge, 1995
- VIZZINISI, E., *Meaning and Cognition*, in «La Nuova Critica», 39, 2002