Università degli Studi di Bologna

FACOLTA' DI LETTERE E FILOSOFIA

Dipartimento di Filosofia

Dottorato di Ricerca in Filosofia

Ciclo XIX

L'APPROCCIO SUBCOGNITIVO ALLO STUDIO DELLA MENTE: MODELLI, CONCETTI, ANALOGIE

Dissertazione di dottorato

Coordinatore
Ch.mo Prof. WALTER TEGA

Presentata dal dottorando:
FRANCESCO BIANCHINI

Relatore
Ch.mo Prof. MAURIZIO FERRIANI

SETTORE SCIENTIFICO-DISCIPLINARE
M-FIL/02

Anno di presentazione
2007

INDICE

Indice 3

Introduzione 5

Capitolo 1 GLI ARGOMENTI DELLA STANZA 11

1.1 La nozione tripartita di “esperimento” 11

1.2 Il cinese macchinoso di Searle 14

1.3 Turing e la stanza dell’intelligenza 18

1.4 Putnam e il telepate giapponese 22

1.5 Lo spostamento della prospettiva 26

1.6 Le obiezioni alla stanza 30

1.7 Il problema di Searle e il “ciclo di purificazione” dei modelli 37

1.8 Leibniz e il mulino della percezione 40

1.9 La stanza fra linguaggio e percezione 44

Capitolo 2 L’APPROCCIO SUBCOGNITIVO ALL’INTELLIGENZA ARTIFICIALE 49

2.1 I principi della subcognizione 49

2.2 La percezione come analogia 52

2.3 L’intelligenza artificiale e il ragionamento analogico 62

2.3.1 Modelli simbolici 64

2.3.2 Modelli connessionisti 68

2.3.3 Modelli ibridi 69

2.4 La questione dei microdomini 71

2.5 L’architettura cognitiva dei modelli 76

2.5.1 I modelli HERSAY e la percezione distribuita del discorso 82

2.5.2 La scansione parallela a schiera 86

Capitolo 3 I MODELLI SUBCOGNITIVI DELLA PERCEZIONE ANALOGICA 95

3.1 Una possibile classificazione 95

3.2 La proposta di un modello teorico 97

3.3 L’alfabeto come universo 103

3.3.1 Il progetto COPYCAT 103

3.3.3 METACAT e i suoi prolegomeni 117
3.4 Il mondo dei numeri in successione 125
3.4.1 SEEK-WHENCE e gli schemi numerici 125
3.4.2 SEQSEE e le nuove strategie auto-osservative 134
3.4.3 SEEK WELL: la matematica come musica 138
3.5 Il mondo reale a tavolino 140
3.6 Frammenti di alfabeti e lettere 152
3.6.1 La sfida dello stile 152
3.6.2 Un modello per il riconoscimento categoriale 156
3.6.3 L'architettura complessa del processo creativo 165
3.7 La geometria come problema limite dell'analogia 173

Capitolo 4 SUBCOGNIZIONE, ANALOGIA E SIMBOLI ATTIVI: VERSO UNA NUOVA TEORIA DEI CONCETTI 187
4.1 Uno sguardo retrospettivo 187
4.2 Scienze, scienze della mente e scienze cognitive 189
4.3 Microprocedure e convalida cognitiva 192
4.4 Microprocedure e computazione: il paradigma della creatività 200
4.5 Microprocedure e cervello: la teoria dei simboli attivi 207
4.6 Modelli dei concetti, concetti come analogie 212
4.7 Conclusione ricorsiva 228

Bibliografia 233

INTRODUZIONE

«Al lettore che preferisco, il quale coltiva in segreto tutti i vizi dell'intelligenza contro i quali combatte; al lettore ipocrita, mio simile e fratello, offro qualcosa su cui meditare»[1], sono le parole, ammiccanti, con cui comincia un lungo e denso studio sul tema che riguarda da vicino quello di questo scritto: il concetto di analogia. Lì, l'autore, Enzo Melandri, delineava lentamente e in maniera certosina, le mille sfumature del concetto nella storia del pensiero e delle teorie che su di esso sono state costruite. Qui, in questo lavoro, molto di quello che viene detto, per quanto posto su un altro piano e affacciato su un altro universo contestuale, risente dell'influenza della storia che della filosofia si può fare attraverso il concetto di analogia, andando a vedere il modo in cui ancora oggi il ragionamento analogico è al centro di un dibattito, di studi e di ricerche nient'affatto secondarie. Quella era un'opera di filosofia. Questa anche, ma in un senso un po' diverso.

L'interesse verso le idee trattate in questa dissertazione ha un duplice fattore motivante: da una parte, certamente, la frequentazione di testi filosofici; dall'altra, la scoperta che un'attenzione sempre maggiore una parte della filosofia ha rivolto a tematiche, da sempre considerate di suo dominio esclusivo, le quali tuttavia negli ultimi decenni sono divenute oggetto di uno studio di tipo scientifico. Stiamo parlando delle ricerche intorno alla mente, al pensiero, al linguaggio e ai concetti, e naturalmente alludiamo al complesso campo di studi che va sotto il nome di scienze cognitive, in cui rientrano le neuroscienze, l'intelligenza artificiale, l'antropologia, la linguistica, la psicologia. Da questa lista non può essere esclusa la filosofia, sia per le ragioni riguardanti l'oggetto della ricerca esposte poc'anzi, sia per la sua, si potrebbe dire, tenacia nel non lasciarsi sfuggire la possibilità di parlare ancora, con pieno diritto, di temi riguardanti la conoscenza e le forme di pensiero. Il prezzo principale che la filosofia ha dovuto pagare nel suo vedersi affiancata da altre discipline in un contesto molto complesso e sfaccettato di indagine è stato quello di doversi abituare al dialogo con le altre "scienze" della mente, un dialogo condotto molto spesso sul loro stesso piano e basato sulla comprensione delle loro metodologie, dei loro principi, dei loro risultati e del linguaggio con cui sono espresse le loro teorie.

Tuttavia, ridurre il ruolo della filosofia nei confronti delle scienze cognitive soltanto a quello di filosofia della scienza, ancorché buona filosofia della scienza, significa lasciarne fuori una gran

[1] Si veda Melandri (2004, p. 3)

parte, trascurarne i suoi apporti, relegarla in uno stato di impasse, dal qual non bastano a smuoverla i suoi intenti di critica e di chiarificazione. La filosofia nei confronti delle altre discipline scientifiche considera di aver svolto il suo dovere e la sua funzione, quando, capendone il linguaggio specifico, riesce a configurarsi come giusta riflessione sul pensiero scientifico in generale e sui principi della scienze in particolare. Questo vale per la filosofia della fisica, della chimica, della biologia, della medicina, della matematica e per l'epistemologia da un punto di vista generale. La filosofia, però, non è solo discussione di teorie. Essa ne è anche artefice, attraverso la sua *vis creativa* e il rigore delle sue argomentazioni. Ciò è vero in particolar modo per le scienze cognitive. Le domande a cui può tentare di dare una risposta sono quelle relative al rapporto mente-corpo o mente-cervello o fra attività cognitive e loro supporto; o anche domande riguardanti la conoscenza, il ragionamento e la "logica" della mente, nel senso di teoria generale dei processi del pensare; o ancora domande sul linguaggio e le sue implicazioni, sul pensiero e i suoi contenuti, sui concetti e i comportamenti che ne derivano, sull'uomo e il suo agire *intelligente* in una realtà sociale in cui gli oggetti reali sono soltanto una parte degli oggetti materialmente individuati e in cui le macchine, le applicazioni tecnologiche, le potenzialità del *software* costituiscono lo sfondo inconsapevole della sua attività quotidiana, eccezionale e scientifica. E sulla quale entro un lasso di tempo presumibilmente breve è atteso l'impatto massiccio della robotica.

In questa prospettiva va considerato il presente lavoro. Esso è certamente un'opera di filosofia della scienza, nel senso che riflette su una serie di teorie e pratiche scientifiche che fanno parte delle scienze cognitive, in particolare teorie e pratiche che rientrano, anche se non esclusivamente, nel campo di studi dell'intelligenza artificiale. Allo stesso tempo questo lavoro intende anche essere una riflessione filosofica su alcuni temi discussi nel dibattito contemporaneo in merito alla natura della conoscenza, della percezione, dei concetti e dell'intelligenza, un dibattito del quale non possono sfuggire le innumerevoli radici nella storia della filosofia. Da questo punto di vista, i riferimenti possibili con i filosofi e le idee del passato sono moltissimi, largamente eccedenti gli scopi di questa trattazione e, dunque, soltanto accennati, a volte in maniera esplicita e in altri casi lasciati sullo sfondo. Non sarà difficile scorgere in alcune riflessioni le influenze del pensiero aristotelico e di quello leibniziano, l'apporto delle idee kantiane e wittgensteiniane, la diffusa pervasività lungo tutto l'arco della dissertazione sia di una concezione pragmatista della realtà sia delle riflessioni di William James sulla "corrente di pensiero"

Obiettivo di questo lavoro è, dunque, anche quello di mostrare che esiste un ponte fra filosofia e scienze cognitive e che l'apporto della prima non è solo quello di una riflessione sui principi o di un'arroccata difesa dei temi di sua stretta competenza, bensì quello di considerarli alla luce dei risultati raggiunti dalle altre discipline che, con un diverso metodo e con altri linguaggi, si occupano delle stesse questioni, rigettando in tal modo le polemiche e le critiche che ciclicamente vengono riproposte da chi si auto-definisce "non filosofo" nei confronti della filosofia. In particolare, le metodologie simulative costituiscono l'opportunità di una banco di prova anche per le teorie

filosofiche più o meno recenti in merito a tutto ciò che, in senso lato, può essere fatto rientrare nell'ambito degli studi sul mentale. Parafrasando un vecchio motto, una delle linee guida del tipo di ricerca compiuto può essere riassunta nella seguente massima:

l'intelligenza artificiale è la continuazione della filosofia con altri mezzi

la quale intende esprimere la profonda compenetrazione che si è avuta nel corso degli ultimi decenni fra riflessioni filosofiche su mente, coscienza, linguaggio e percezione, e la progettazione di sistemi e programmi per il calcolatore – per tipi di calcolatore sempre più potenti dal punto di vista delle risorse computazionali – con il fine di comprendere meglio il fenomeno dell'"intelligenza" e il pensiero umano. È noto che i confini dell'utilizzo a fini di ricerca delle tecniche simulative si sono allargati, negli ultimi tempi, allo studio di tutto ciò che può essere fatto rientrare all'interno dell'"orizzonte cognitivo", fino a includere il modo in cui fenomeni di questo tipo possono essere attribuiti al mondo animale, la comprensione di come tali fenomeni si siano prodotti, e *soprattutto* si possano produrre attraverso dinamiche evolutive, e lo studio della struttura dei sistemi complessi in grado di esibire un comportamento che dall'esterno viene considerato "intelligente". Con questo termine si vuole intendere un comportamento nel quale esiste un *gap*, una frattura, fra le condizioni iniziali e l'obiettivo finale, una frattura che è al tempo stesso lacuna esplicativa, luogo nascosto dei meccanismi processuali, superamento sia del vincolo controintuitivo della dicotomia stimolo-risposta sia di un'interpretazione troppo semplicistica delle leggi metafisiche di azione-reazione e di causa-effetto.

Le tesi principali sostenute in questo lavoro sono tre. La prima è che un modo proficuo di indagare i fenomeni cognitivi è quello di porsi a un livello intermedio fra processi mentali superiori e attività cerebrale. Tale assunto ha come ricaduta metodologica lo studio per via simulativa dei processi di pensiero attraverso lo sviluppo di sistemi che si situano all'interno del paradigma della complessità e le cui architetture modulari sono basate sullo scambio interattivo di informazione in un ciclo dinamico di avvicinamento alla "soluzione", cioè alla produzione di una risultato al termine dell'esecuzione della prestazione. Tali sistemi sfruttano euristiche basate sull'elaborazione stocastica e parallela, si avvalgono di opportune funzioni di auto-controllo e monitoraggio della propria attività, fanno uso di una certa quantità di elementi casuali e allo stesso tempo sono vincolate da attrattori che stabilizzano la dinamica dell'elaborazione verso processi deterministici risultati ben definiti. Tutto ciò è permesso dall'interazione di micro-agenti che rispecchiano nella simulazione i processi che ricadono al di sotto della soglia dell'attività cosciente del pensiero. Essi permettono la modellizzazione dei processi di percezione di alto livello, cioè quelli in cui si fondono le conoscenze già possedute e gli input esterni, e i processi, strettamente intrecciati ai primi, di creazione di analogie.

La seconda tesi è l'affermazione secondo la quale un ruolo centrale nei processi di pensiero è ricoperto dalla rappresentazione e dalle modalità rappresentative di ogni sistema che si vuole definire come intelligente. Perciò, i limiti dell'approccio tradizionale allo studio del ragionamento, basato sulla creazione di sistemi in cui la conoscenza è espressa in forma simbolica e l'elaborazione assume il carattere di una derivazione inferenziale a partire da sistemi di credenze espresse nel calcolo dei predicati, non porta ad un rifiuto della rappresentazione come elemento fondamentale dei processi di pensiero, bensì ad un suo adattamento in un contesto come quello dei sistemi descritti in precedenza in cui devono trovare posto *anche* le forme della logica. La tesi consiste, dunque, nel non respingere *in toto* la nozione di, e la funzione della, rappresentazione, ma nella loro rivisitazione; nel non rifiutare il simbolico, ma nell'accettare che alcune funzioni cognitive, in particolare i processi di interazione con una ambiente esterno e con un input variabile, richiedono il dispiegarsi di un *simbolismo statistico, strutturale e a soglia* che permetta ad un sistema la costruzione di adeguate rappresentazioni della realtà esterna e della propria attività in essa.

La terza tesi, corollario delle prime due, è che dal punto di vista della conoscenza, il livello intermedio *e* simbolico è quello dei concetti, elementi *rappresentativi-attivi*, strettamente correlati ma distinti dal linguaggio che li esprime. L'indagine in merito alla loro natura, condotta attraverso l'implementazione di modelli simulativi, e tenendo anche conto allo stesso tempo degli apporti teorici della filosofia e della psicologia, costituisce una *conditio sine qua non* della comprensione dei processi di pensiero, la quale risulta verosimilmente soddisfatta nel momento in cui si possa dare una teoria unificata dei concetti che spieghi tutti i fenomeni connessi con i concetti e la concettualizzazione. Una teoria di questo tipo, ancora *in fieri* e passibile di ulteriori approfondimenti, è quella che viene proposta alla fine di questo lavoro e che deriva direttamente dalle ricerche simulative compiute all'interno dell'approccio intermedio allo studio dei processi di pensiero definito, proprio per questa ragione, *subcognitivo*. Chiamo questa teoria: la teoria dei concetti come analogie.

La ricerca che ha portato alla stesura della dissertazione ha avuto diverse fasi, dall'approfondimento delle tematiche di filosofia della mente, della scienza e del linguaggio connesse a questi temi, all'interazione con i programmi che implementano la prospettiva simulativa considerata, sviluppati da Douglas Hofstadter e dal *Fluid Analogies Research Group* negli ultimi anni principalmente presso l'*Indiana University* di Bloomington. Di tali sistemi viene presentata un'analisi dettagliata in una prospettiva conforme alle tesi sostenute in questo lavoro, che ne mostra l'evoluzione nel corso di quasi tre decenni, evoluzione che rispecchia in larga parte quella delle scienze cognitive, sia dal punto di vista teorico che pratico.

La dissertazione si compone di quattro capitoli. Il primo delinea il contesto teorico nel quale sono nate le critiche di natura filosofica all'intelligenza artificiale tradizionale, critiche che attraverso una serie di *Gedankenexperiment* e di argomentazioni hanno mirato allo svuotamento dell'impostazione "linguistico-simbolica" allo studio dei processi di pensiero. Il capitolo si conclude con

l'affermazione del ruolo centrale della percezione nei processi di pensiero, un pensiero che va considerato, perciò, all'interno di un *contesto di rappresentazione*, che rispecchia, fra le altre cose, un corpo, un ambiente e una rete sociale di attori cognitivi. Nel secondo capitolo vengono discussi gli aspetti fondamentali dell'approccio definito subcognitivo, a partire dalle prime considerazioni del promotore di tale approccio, Douglas Hofstadter, le idee del quale costituiscono uno dei riferimenti principali di tutta la trattazione. Inoltre, in esso è affrontato anche il tema dell'analogia dal punto di vista cognitivo e vengono discussi i principali modelli simulativi dedicati al ragionamento analogico. Nel terzo capitolo trova ampio spazio l'analisi dettagliata dei modelli subcognitivi, considerati nella prospettiva del dominio di applicazione, attraverso la quale è possibile constatare l'evoluzione dei modelli verso un sempre maggiore arricchimento dell'architettura e della rappresentazione della conoscenza che hanno l'obiettivo di catturare, e muoversi in, domini sempre più complessi. Infine, nel quarto capitolo si procede ad una discussione generale dei modelli e dell'impostazione di ricerca ad essi connessi, ma allo stesso tempo delle teorie sul mentale e sui concetti di cui essi intendono essere una realizzazione effettiva.

Il capitolo conclusivo, come molte volte accade alla fine di una ricerca, soprattutto se inserita in un campo di studi così complesso, non giunge a risultati assoluti, ma ad esiti suscettibili di ulteriori approfondimenti e investigazioni, e a un epilogo su mente, cervello e concetti quasi aporetico se, per ritornare circolarmente agli inizi di questa breve introduzione, non si tiene conto della necessarietà di un continuo dialogo epistemologico in merito ai temi affrontati. Nelle scienze cognitive, infatti, più ancora che nelle altre scienze, la riflessione sui principi, sul linguaggio e sui concetti utilizzati ha un peso così notevole, da impedire, qualora manchi, il loro fruttuoso svolgimento, l'acquisizione di risultati riconosciuti come certi e condivisi, e la funzione esplicativa e predittiva della realtà che ogni sapere scientifico brama.

Ringraziamenti

In un tempo in cui svolgere attività di ricerca è diventato un impegno nel quale le difficoltà esteriori soverchiano il già pur difficile compito di muoversi nel territorio dell'inesplorato i ringraziamenti assumono un valore contestuale superiore a quello di semplice moto interiore.

Portare a termine questo lavoro non sarebbe stato possibile, innanzitutto, senza l'aiuto morale e materiale della mia famiglia e dei miei genitori in particolare.

Ringrazio, inoltre, le interessanti intelligenze che hanno guidato la mia ricerca a cominciare da Maurizio Matteuzzi, che si è pazientemente sobbarcato l'oneroso compito di leggere l'intero lavoro e ha messo a mia disposizione le sue conoscenze; Douglas Hofstadter, con il quale ho potuto discutere in innumerevoli conversazioni le idee e le tesi esposte in questa opera e che ha permesso che io svolgessi una parte del periodo di dottorato presso il *Center for Research on Concepts and*

Cognition (CRCC) dell'*Indiana University* a Bloomington; Roberto Cordeschi, che a più riprese mi ha fornito utilissimi consigli sul modo in cui impostare l'intera ricerca; Giorgio Sandri, al quale sono debitore, tra le altre cose, di molte illuminanti indicazioni sui temi della computazione e dei sistemi automatici sia da un punto di vista logico che filosofico; i membri del *Fluid Analogies Research Group* con cui ho potuto entrare in contatto e discutere gli aspetti tecnici e teorici dei modelli presentati in questo lavoro: Abhijit Mahabal, Francisco Lara-Dummer, Eric Nichols, Damien Sullivan e Matt Rowe.

Un ringraziamento va anche a chi ha supportato questo lavoro dal punto di vista pratico: Helga Keller, che, in qualità di amministratrice del CRCC, ha organizzato entrambi i miei soggiorni americani; Cristina Paoletti, impagabile nel risolvere tutti i dubbi e i problemi burocratici sorti in questi tre anni di dottorato; il Dirigente dell'Istituto Statale di Istruzione Superiore "Archimede" di San Giovanni in Persiceto, Giuseppe Riccardi, e la Segreteria dell'Istituto, per avermi consentito di svolgere a tempo pieno il lavoro di ricerca nell'ultimo anno di dottorato.

Ringrazio ancora, Raffaela Serrani per le sue consulenze linguistiche sulle traduzioni da me effettuate e per il paziente lavoro di supervisione delle mie idee; Giuliano Bettella e Viola Bertazzini, per le numerose discussioni sui temi affrontati in questi capitoli e sul senso della ricerca in filosofia e nelle scienze cognitive; Alfio Gliozzo per i chiarimenti su questioni di linguistica computazionale; Elisabetta Versace, per avermi dato l'opportunità di discutere parte di queste idee in più occasioni con altri studiosi e ricercatori all'Università di Trieste.

Ringrazio, infine, tutti i componenti del Progetto M. per le molto acute conversazioni sugli aspetti più profondi e interiori del mestiere di ricercatore, che hanno condiviso con me nel corso di questi anni.

Un pensiero, oltre che un ringraziamento, va a Maurizio Ferriani, che ha visto gli inizi e purtroppo non la fine di questa ricerca, guidandone i primi passi e orientandone verso un faro filosofico il suo senso complessivo. A lui questo lavoro è dedicato.

Avvertenza

Tutte le citazioni nel testo sono in italiano. Per esse, dove non diversamente specificato, si è fatto ricorso alle traduzioni italiane disponibili e segnalate in bibliografia. Il riferimento all'opera, indicato con il sistema autore-anno, riporta la data di pubblicazione originale dell'opera, ma la pagina o le pagine della citazione sono quelle della traduzione italiana, se presente. In tutti gli altri casi l'autore delle traduzioni è il medesimo di questo lavoro, delle quali pertanto si assume ogni responsabilità.

Capitolo 1

GLI ARGOMENTI DELLA STANZA

1.1 La nozione tripartita di “esperimento”

Una parte essenziale della ricerca scientifica consiste nella sperimentazione. Un esperimento serve a comprovare o a invalidare una particolare teoria attraverso la conferma o meno delle previsioni compiute in base ad essa. Sebbene ogni disciplina scientifica abbia la sua particolare metodologia sperimentale questo schema generale è condiviso da tutte le scienze particolari: vengono fatte delle ipotesi; si procede alla ricerca sul campo o si costruisce in laboratorio una situazione in cui tali ipotesi possano essere messe alla prova; si confrontano i risultati ottenuti con le ipotesi iniziali per verificare il grado di esattezza delle previsioni.

Esiste, tuttavia, un altro impiego legittimo del termine “esperimento”, anche se del tutto differente, il quale si riferisce alla ideazione di situazioni puramente teoriche non passibili, in senso contingente o assoluto, di una effettiva realizzabilità pratica: gli esperimenti mentali. Va da sé che compiere un esperimento scientifico attenendosi a una metodologia precisa e riproducibile è molto diverso dall’ipotizzare una situazione ideale in cui viene messa alla prova l’efficacia di alcuni concetti nel descrivere situazioni teoricamente concepibili, o, in altri termini, nell’inscenare mondi possibili. Tuttavia, l’utilità di un esperimento mentale è indubbia. Attraverso di esso circostanze empiricamente irrealizzabili, sia nel senso di una mancata acquisizione tecnica meramente contingente, come nell’esperimento einsteiniano dei gemelli in merito alla relatività delle misurazioni temporali (viaggi su scala macroscopica a una velocità confrontabile a quella della luce potranno essere resi disponibili dallo sviluppo di nuove tecnologie), sia in quello di un’impossibilità fattuale non vincolata temporalmente, come nel caso della nave galileiana utilizzata per illustrare la relatività del moto (nel mondo reale non si darà mai il caso di un moto costantemente uniforme), possono essere costruite e logicamente testate per verificare la plausibilità delle formulazione teoriche cui assegniamo il compito di spiegare scientificamente, cioè in modo prevedibile, la realtà. Fra gli esperimenti mentali possono essere annoverati, ad esempio, stati di cose controfattuali, ma anche situazioni che descrivono fenomeni difficilmente esperibili dal punto di vista empirico, come nel caso del problema della definizione dell’identità personale e del suo legame con il suo substrato materiale nell’individuo cui viene riconosciuta, il cervello; per ragioni metodologiche, ma anche per

ovvie ragioni etiche, l'impostazione di un profilo di indagine volta a chiarire le questioni dell'"Io" e della "Coscienza del sé" sembrano trarre vantaggio dall'utilizzo, non esclusivo[1] certamente, di *Gedankenexperimente*[2].

Queste due modalità di esperimento non costituiscono, perciò, una contrapposizione metodologica all'interno della pratica scientifica, ma due diversi approcci complementari attraverso cui il sapere procede volti a testare in maniera molto diversa la nostra generale concezione della realtà. Se l'esperimento scientifico ha come fine quello di indagare un qualche aspetto del mondo fenomenico, l'esperimento mentale si pone l'obiettivo di vagliare la verosimiglianza delle teorie e la coerenza logica delle assunzioni fondamentali sulle quali esse si reggono, anche per mezzo della costruzione di situazioni contrarie all'intuizione allo scopo di indirizzare la ricerca scientifica verso un cammino piuttosto che un altro, fatto salvo il vincolo della loro confrontabilità con la realtà *almeno* sotto certi aspetti; l'uno ha la dalla sua la forza del fatto, l'altro la seduzione del concepibile.

Non è raro che la molla che faccia scattare un'ampia serie di discussioni sia proprio un esperimento mentale, anche se la sua peculiarità non è quella di inscenare e/o indagare una situazione reale, nella quale vengono messe alla prova determinate proprietà di entità che prendono parte a qualche fenomeno, ma quella di descrivere uno stato ideale di cose, una narrazione che getti luce sulla plausibilità o non plausibilità dell'utilizzo di determinate categorie e di una particolare teoria per spiegare specifici effetti. Le discussioni che seguono la formulazione di un esperimento ideale permettono di ricostruire il succedersi delle teorie proposte per la spiegazione di determinati fenomeni all'interno di una particolare disciplina scientifica, o di un programma di ricerca che intenda spiegare alcuni aspetti della realtà, i quali costituiscono l'obiettivo di differenti ambiti scientifici. Tale ricostruzione permette la valutazione dei principi epistemologici e delle assunzioni fondamentali che regolano le teorie, la loro revisione, cancellazione o integrazione con nuove ipotesi e principi, anche se non la loro validazione empirica.

Ho utilizzato la dicitura di "argomenti della stanza" per riferirmi ad alcuni esperimenti ideali che non "provano" nulla nel senso usuale, scientifico del termine, ma che aiutano a riflettere sul senso di una teoria, che in questo caso è psicologica, e sui principi epistemologici in base ai quali è costruita nel tentativo di fornire una spiegazione verosimile dei fenomeni mentali, delle attività cognitive e dei processi di pensiero. La stanza in questione è un luogo metaforico, che indica i limiti esterni di ciò che viene studiato e che deve essere compreso e spiegato. Entrare nella stanza, o esservi dentro, costituisce la mossa che dà l'avvio alla formulazione di ipotesi esplicative. Ciò che

[1] Dal punto di vista neurofisiologico, ad esempio, un utile apporto può venire dall'impiego a fini sperimentali della risonanza magnetica funzionale. Per quanto riguarda il versante psicologico, valgano come esempio tipico gli esperimenti di Gallup con gli scimpanzé sul riconoscimento dell'identità. Per questi si rimanda a Gallup (1970).

[2] Sul modo in cui gli esperimenti mentali possono essere utilizzati per studiare il fenomeno della coscienza si veda, ad esempio, Robinson (2004).

viene ipotizzato sono i contenuti della stanza, esattamente nello stesso modo in cui sono stati ipotizzati nel corso di un cinquantennio (o di cinque secoli o di più di due millenni) i contenuti spirituali, materiali, organizzativi e strutturali, della mente e, in tempi più recenti, del cervello.

Anticipo fin da queste prime pagine una proposta metodologica. Alle due tipologie di esperimento appena menzionate se ne può aggiungere una terza, peculiare e caratterizzante le scienze cognitive: quella relativa alla realizzazione effettiva di simulazioni dei fenomeni mentali, nel senso più ampio del termine "simulazione", che coinvolge le differenti impostazioni delle scienze cognitive, quella più tradizionale e quella cosiddetta nuova (prima connessionista e poi evolutiva), ma anche ogni prospettiva che preveda il superamento della loro contrapposizione a favore di una più produttiva complementarietà.

Se tale superamento è ciò che sta perlopiù avvenendo, sia sul piano retorico che su quello che effettivo, nei programmi di ricerca dedicati allo studio dei fenomeni mentali, appare necessario un riesame della nozione di simulazione, un suo affinamento e perfezionamento, strettamente legato alla rapida trasformazione delle discipline coinvolte nelle scienze cognitive. La mia proposta, in questa fase iniziale, è quella di considerare gli *esperimenti simulativi* una sorta di grado intermedio fra gli esperimenti scientifici tradizionali e quelli mentali. Dei primi dovrebbero ereditare la prevedibilità, la riproducibilità, la ricerca dell'esattezza quantitativa, laddove possibile, l'intersoggettività e la governabilità; dei secondi la plausibilità e la coerenza logica, o la loro esplicita e consapevole negazione (come nei procedimenti per assurdo e nei controfattuali), la libertà dell'intuizione che di volta in volta li ispira e il grado di realismo che li rende accettabili e confrontabili con il fenomeno che deve essere indagato. Un esperimento mentale è di una qualche utilità, infatti, se la situazione che in esso viene descritta conserva collegamenti diretti ed espliciti con la realtà, soprattutto nel caso di fenomeni la cui diretta osservabilità nel mondo reale risulta alquanto problematica da definire. Il riferimento alla realtà è un tratto di cui ogni simulazione cognitiva deve *a fortiori* tenere conto, un vincolo cui non può non soggiacere.

Questo è l'inizio. La tripartizione suggerita verrà ripresa nella parte finale di questo scritto. Ora intendo esporre e valutare una serie di *Gedankenexperiment* che hanno avuto una grande influenza sullo studio dei processi cognitivi. Le riflessioni che ne scaturiranno saranno la base per ulteriori considerazioni in merito a un particolare approccio alle discipline simulative, che possiamo definire in via provvisoria e concordemente con i suoi ideatori "subcognitivo", e che è un tentativo di spiegazione di molteplici fenomeni connessi alla produttività e alla creatività del pensiero.

1.2 Il cinese macchinoso di Searle

Nel settembre del 1980 usciva sulla rivista *The Behavioral and Brain Sciences* il noto articolo di Searle "Menti, cervelli e programmi"[3] insieme alle obiezioni da parte di un discreto numero di interlocutori appartenenti a diverse estrazioni disciplinari (filosofi, psicologi, scienziati cognitivi ed esperti di intelligenza artificiale) e alle conseguenti articolate risposte dello stesso Searle. Il dibattito che ne scaturisce negli anni seguenti e che fa eco, allargandolo, a quello iniziale contenuto nella pubblicazione, trae la sua linfa più che ragionevolmente dalle forti suggestioni provocate dal nucleo centrale dell'articolo, il *Gedankenexperiment* della stanza cinese. Infatti, pur prefiggendosi Searle il duplice scopo di circoscrivere la possibilità dell'intenzionalità a ogni meccanismo che avesse proprio gli stessi poteri causali del cervello e di dimostrare che tale possibilità non rientrasse in alcun modo e a nessuna condizione tra le caratteristiche ascrivibili a un sistema computazionale artificiale, finì per destare, introducendo l'argomento della stanza cinese, un'attenzione ben maggiore a quella semplicemente riservata alla valutazione della validità delle conclusioni da lui raggiunte nell'articolo.

L'intento polemico è diretto fin dal principio, per esplicita dichiarazione dell'autore, a quella che viene definita Intelligenza Artificiale (d'ora in avanti IA) forte e che è caratterizzata, secondo Searle, dalla seguente ipotesi alla base del suo programma di ricerca:

> Il computer appropriatamente programmato è *realmente* una mente, nel senso che i computer, cui sono stati dati i programmi giusti, *capiscono* e hanno altri stati cognitivi. Nella IA forte, per il fatto che il computer programmato ha stati cognitivi, i programmi non sono semplici strumenti che ci rendono possibile considerare spiegazioni psicologiche: piuttosto i programmi costituiscono di per sé le spiegazioni. (Searle, 1980, 46)

La dimostrazione della palese assurdità di tali affermazioni dovrebbe giustificare, secondo l'autore, l'abbandono di una concezione forte dell'IA, basata sulla manipolazione formale di simboli, che identifica programmi e processi di pensiero, a favore di una concezione debole, in cui l'utilizzo del calcolatore è soltanto uno strumento ausiliario e non sostanziale, seppur definito «molto potente» (Searle, 1980, p. 46), per la comprensione e per la spiegazione delle attività cognitive. A questa presa di posizione critica nei confronti di una certa visione, a grana larga, dell'IA Searle fa seguire la parte *construens* del suo argomento, la tesi secondo cui, data l'incapacità di un sistema computazionale artificiale di essere intenzionale, soltanto il cervello, attraverso i suoi poteri causali, è in grado di produrre l'intenzionalità. Ne consegue che non è lecito affermare che non esistono macchine pensanti, poiché il cervello stesso è una macchina, però è una

[3] Il saggio, dal titolo «Mind, Brains and Programs», compare per la prima volta su *The Behavioral and Brain Sciences*, nel 1980. La prima edizione italiana è nel volume *Menti, cervelli e programmi, un dibattito sull'intelligenza artificiale*, a cura di Graziella Tonfoni, nel quale è anche riportata per intero la serie delle obiezioni e la risposta dell'autore.

macchina di tipo speciale, dotata di particolari poteri che derivano dalla sua struttura biochimica. Solo una macchina «con nessi causali interni che sono equivalenti a quelli dei cervelli» (*ibidem*) può foggiarsi della qualifica di macchina pensante, o, detto in altri termini, può dirsi dotata di intenzionalità.

Con il termine "intenzionalità", uno dei più significativi e discussi nella filosofia del Novecento[4], Searle intende qualcosa di intrinsecamente connesso con il cervello e con ogni tipo di macchina che presenti la stessa struttura causale di quella del cervello. Ciò non equivale a escludere che altre entità, rispetto agli individui umani, possano avere la capacità di pensare, ma è, di fatto, una restrizione di livello molto elevato, non unicamente perché soltanto gli umani sembrano dotati di tali poteri casuali legati al cervello, ma anche perché, nel corso del suo saggio, Searle non specifica affatto in cosa consistano questi poteri[5]. Egli si limita ad affermare che «qualunque cosa faccia il cervello per produrre intenzionalità, questa non può consistere nell'istanziare un programma, poiché nessun programma, di per sé è sufficiente per l'intenzionalità» (Searle, 1980, p. 72), il che equivale a dire che l'intenzionalità, considerata come la caratteristica qualificante del mentale, non può essere ottenuta attraverso il mero, formale, inumano potere computazionale dei calcolatori.

Per meglio definire il suo attacco al computazionalismo, Searle non circoscrive la sua argomentazione ai calcolatori. Anzi, conferisce forza alla sua argomentazione, sostenendo che anche la mente che si comporta in maniera computazionale, cioè che manipola formalmente dei simboli, è priva del potere intenzionale. Queste affermazioni, presenti a più riprese nell'articolo, corroborano l'idea che Searle non sia un teorico dell'anti-meccanicismo, che la sua non sia una presa di posizione contro l'intelligenza artificiale *tout court*. Al contrario, proprio la sua insistenza sui poteri causali del cervello denoterebbe il suo favore nei confronti di un'interpretazione dell'IA di stampo connessionistico, se questo filone di indagine allo studio dei processi di pensiero fosse ancora non pienamente tornato alla ribalta ai tempi in cui egli scrive. La vecchia IA simbolica appare un costrutto teorico sul punto di saltare definitivamente dopo aver visto il suo sviluppo, i suoi successi e i suoi fallimenti, cioè dopo essere passata, nel corso degli anni sessanta e settanta del Novecento, attraverso la riconsiderazione realistica dei tempi di raggiungimento dei suoi obiettivi, nel corso della quale vengono ridimensionate le previsioni e le aspettative entusiastiche suggerite dalla proto-intelligenza artificiale degli anni cinquanta.

Tuttavia, ricondurre l'essenza del mentale (quale è l'intenzionalità per Searle, in un senso che si potrebbe definire deittico) ai poteri causali del cervello e negare qualunque pregnanza cognitiva alla manipolazione formale di simboli non ci dice alcunché in merito a questi poteri, né garantisce il fallimento del computazionalismo e la fallacia di ogni argomentazione in appoggio all'idea della riproduzione delle attività mentali attraverso l'implementazione di programmi che operano su simboli.

---

[4] Per una rassegna storica delle teorie delle intenzionalità si rimanda a Gozzano (1997). Per un affondo nel dibattito in fiosofia della mente che qui ci interessa si veda anche Dennett (1989).

[5] Questa è proprio una delle principali obiezioni che vengono rivolte a Searle dai suoi interlocutori nelle *replies*.

Per dare forza alle sue tesi egli si serve dell'esperimento della stanza cinese, che costituisce il nucleo centrale del suo saggio e il perno attorno cui costruisce le sei possibili obiezioni alle sue affermazioni. L'argomento è noto e trae la sua forza dalla plausibilità intuitiva degli elementi su cui è costruito. Qui ne viene riportata una versione lievemente modificata rispetto a quella che presenta Searle, nel senso che la struttura dell'esperimento è anteposta in modo che risalti all'esposizione del fenomeno che deve essere spiegato, cioè la comprensione di una situazione espressa in linguaggio naturale.

In breve e in maniera semplificata, l'esperimento mentale consiste nell'immaginare un individuo di madrelingua inglese e completamente ignorante della lingua cinese chiuso in una stanza e intento a compiere operazioni servendosi dell'ausilio di un manuale di regole scritte in inglese. Tali regole permettono all'individuo di comunicare all'esterno della stanza attraverso un interfaccia una serie di caratteri cinesi che sono *in una qualche relazione* con ideogrammi cinesi inviati nella stanza attraverso l'interfaccia da agenti esterni. Questi possono essere indifferentemente individui di madrelingua cinese o programmatori a conoscenza delle stesse regole dell'individuo all'interno della stanza. Quello che conta è ciò avviene *dentro* la stanza e come viene interpretato all'esterno. L'uomo nella stanza si trova nella situazione di possedere due plichi di fogli contenenti scritture cinesi e attraverso l'uso delle regole (in inglese) del manuale è in grado di correlare i simboli cinesi che gli vengono consegnati dall'esterno con i simboli cinesi dei plichi in suo possesso al fine di rendere ai suoi interlocutori esterni una serie di simboli cinesi attraverso l'uso esclusivo delle regole contenute nel manuale. Tale manuale consiste, in definitiva, di una serie di istruzioni alla stregua di un programma, anzi, di un insieme di programmi. Per mezzo di tali istruzioni è possibile «mettere in relazione una serie di simboli formali con un'altra serie di simboli formali (e tutto quello che formale significa qui, è che posso identificare i simboli interamente attraverso le loro forme)» (Searle, 1980, p. 48). Le regole permettono l'istituzione di correlazioni fra simboli esclusivamente in base alla loro forma attraverso un processo che deve essere di questo genere: prendo in considerazione un simbolo o un insieme di simboli cinesi fra quelli che mi vengono dati; lo cerco sul manuale; vedo a quale simbolo o insieme di simboli cinesi corrisponde nei plichi; leggo le istruzioni che devo attuare una volta instaurata questa correlazione; riproduco alcune forme, cioè scrivo nuovi simboli in cinese, basandomi sulle istruzioni che caratterizzano il simbolo o l'insieme di simboli correlati, giacché per il tramite delle regole io sono in grado di «riprodurre certi simboli cinesi con certi tipi di forme datemi» (*ibidem*) dagli interlocutori fuori della stanza; infine, invio queste nuove forme – ideogrammi cinesi – all'esterno attraverso l'interfaccia.

Presentato in questo modo l'argomento sembra procedere senza intoppi. Searle conta sul fatto che continui a funzionare anche quando si pensi ad esso come alla descrizione in termini meccanici di una particolare attività cognitiva di alto livello: la comprensione di narrazioni in linguaggio naturale. Infatti, i plichi di fogli contenenti simboli cinesi in dotazione all'individuo nella stanza corrispondono: il primo, alla situazione prototipica dell'andare a cena in un ristorante (dotata di

caratteristiche standard: arrivare, sedersi, consultare il menù, ordinare, aspettare le portate, mangiare, pagare e andarsene);il secondo, al racconto di un particolare episodio di cena al ristorante (in cui un distinto signore va al ristorante e ordina una bistecca; quando gli viene portata si accorge che è bruciata; allora si alza e se ne va). I simboli cinesi che vengono inviati dall'esterno nella stanza sono domande sulla storia (del tipo: il signore ha mangiato la bistecca? Ha pagato il conto?); quelli che vengono inviati all'esterno sono risposte pertinenti alle domande. Il manuale in inglese di correlazione dei simboli cinesi serve a produrre le risposte e corrisponde a un programma inserito in un calcolatore che, afferma Searle, si presume comprenda la storia, mostrandolo nella pertinenza delle risposte alle domande[6]. Ovviamente l'individuo nella stanza non capisce la storia, né, ugualmente, le domande che gli vengono fatte e le risposte che fornisce. Di conseguenza la costruzione del meccanismo della stanza fa svanire l'intenzionalità; ma anche la *elimina*. Questo fatto viene fatto corrispondere alla sua *mancata spiegazione* in termini computazionali.

La forza dell'argomento risiede nella sua immediatezza. Chi potrebbe affermare di non capire la sua lingua nativa? Chi non negherebbe recisamente la comprensione una lingua che non solo non ha mai appreso, ma che è anche così diversa per quanto riguarda la sua notazione grafica, nonché probabilmente per molteplici aspetti sintattici e semantici? Il fatto che si possa immaginare di istituire una relazione puramente formale fra i simboli delle due lingue in base alla quale possano essere compiute operazioni che danno l'idea a chi sta fuori della stanza di conversare attraverso l'interfaccia con un madrelingua cinese, *mentre non è affatto così*, giustifica, secondo Searle, l'affermazione secondo cui la mera manipolazione formale di simboli, cioè l'istanziazione di un programma, non garantisce di certo la comprensione dei simboli che si stanno manipolando e, a cascata, l'intenzionalità. Detto in altri termini, se io, essere umano, comportandomi come un computer che esegue una serie ordinata di istruzioni non comprendo i simboli cui applico le istruzioni, *a fortiori* non li comprenderà il programma.

Con questo argomento Searle intende portare una critica forte alla ricerca in IA compiuta a Yale da Roger Schank e al suo modello per la comprensione di brevi narrazioni riguardanti episodi specifici[7]. L'intento polemico, peraltro, non era direttamente rivolto all'inadeguatezza nel catturare la complessità del mondo reale da parte di *script*, e *frame*, strutture di rappresentazione delle conoscenza che condividono in buona parte medesimi assunti teorici di fondo. Infatti, se gli *script* e i *frame* servono a catturare situazioni del mondo reale attraverso una maschera informazionale prototipica, che è un modello standard della realtà, cioè di una porzione specifica della realtà, lasciando aperta la possibilità dell'inserimento di dettagli specifici non contemplati nel prototipo,

[6] Non necessariamente attraverso l'esattezza delle risposte. La pertinenza è una nozione maggiormente comprensiva e contempla la possibilità che si diano anche risposte sbagliate, come può accadere a un soggetto umano chiamato a mostrare la sua comprensione della storia presentata. L'importante è che le risposte sbagliate non siano troppo fuori bersaglio. La nozione di pertinenza ha il vantaggio di cogliere la vaghezza, piuttosto che l'assolutezza, della nozione di comprensione.

[7] Searle si riferisce nel suo articolo a Schank e Abelson (1977), il saggio in cui viene esposta la teoria psicologica degli *script* come modelli strutturati della comprensione del mondo reale. Per una esposizione più dettagliata delle tesi di Schank e dei programmi sviluppati dal gruppo di ricerca di Yale si veda Schank (1984).

essi difettano nel cogliere a pieno la complessità imprevedibile e potenzialmente infinita delle situazioni del mondo reale racchiudibili in un copione o in uno schema prestabiliti, nonostante la non-monotonicità con cui trattano l'informazione[8]. Dunque, stando così le cose, quale è il vero bersaglio della critica di Searle? Per rispondere a questa domanda occorre considerare alcuni presumibili retroscena del suo *Gedankenexperiment*.

1.3 Turing e la stanza dell'intelligenza

Facciamo un passo indietro e andiamo a rivedere come Turing introduce il celebre gioco dell'imitazione per valutare la plausibilità di macchine intelligenti. Egli si propone di rispondere a una domanda semplice e diretta: possono le macchine pensare? Per stabilirlo, Turing propone un gioco che consiste nel considerare tre individui: un uomo, una donna, un interrogante, la natura del cui sesso non è importante ai fini del gioco. L'interrogante occupa una stanza e l'uomo e la donna si trovano in un'altra, separata dalla prima e da cui è permessa la comunicazione con l'interrogante soltanto attraverso una telescrivente. Scopo ultimo del gioco per l'interrogante è quello di capire quale dei due individui sia l'uomo e quale la donna. La parte interessante di questo altro *Gedankenexperiment* sta nel chiedersi: che cosa succede se il posto dell'uomo viene preso da una macchina, ovvero da un computer che esegue un particolare programma atto a simulare le capacità umane?

La vasta letteratura di discussione scaturita dalla presentazione del gioco dell'imitazione ha in genere tralasciato di considerare essenziale al gioco la figura della donna nella versione in cui l'uomo viene sostituito da una macchina, reputando, a ragione, che l'indistinguibilità fra essere umano e macchina fosse l'obiettivo primario per la riuscita del gioco e, quindi, per il superamento del test. Tale indistinguibilità, infatti, sarà raggiunta soltanto quando nella separazione di interrogante e macchina, il primo avrà l'impressione di dialogare con la seconda come farebbe con un essere umano, sia esso uomo o donna. La separazione diviene condizione fondamentale per la conduzione del gioco e nella separazione l'unico punto di contatto non può che essere, e non deve essere, altro che una comunicazione di tipo linguistico. Questo per due ragioni.

La prima è che non ci devono essere restrizioni di tipo macro-biologico, intendendo con esse le maggiori o minori capacità attestabili in un individuo dal punto di vista corporeo nei confronti di una macchina e delle sue prestazioni. Infatti, afferma Turing, non sarebbe corretto «penalizzare la macchina per la sua incapacità di brillare in un concorso di bellezza, né penalizzare un uomo perché

[8] La possibilità di inserire dati relativi alla situazione specifica di volta in volta diversi all'interno di una descrizione standard è ciò che rende gli *script* e i *frame* uno dei tentativi più riusciti di superare le limitazioni del calcolo dei predicati come sistema di rappresentazione della conoscenza, limitazioni dovute alla sua monotònicità. Tentativi di introdurre specifiche modificazioni e regole per esprimere formalmente (nel calcolo dei predicati) il ragionamento non-monotòno sono stati compiuti fin dagli anni sessanta del Novecento. Per un'ampia panoramica su tale questione si rimanda a Fisher-Servi (2001).

perde una corsa contro un aeroplano. Le *condizioni* del nostro gioco rendono irrilevanti queste incapacità. [...] l'interrogante non può chiedere dimostrazioni *pratiche*» (Turing, 1950, p. 169. [*enfasi mia*])[9]. Sulla stessa linea sembrerebbero classificabili anche strumenti che realizzano prestazioni definite intelligenti, ma puramente meccaniche, quali, tanto per fare un esempio, un sistema di antibloccaggio dei freni delle ruote di un veicolo (comunemente chiamato ABS). Un apparecchio di tal genere viene utilizzato per migliorare la prestazione umana in fatto di frenata. Utilizza un sistema di retroazione fra la pressione del pedale da parte dell'individuo e l'attrito del fondo stradale al momento della frenata, ed è proprio a causa del miglioramento prodotto rispetto all'uomo che il sistema viene definito intelligente. Tuttavia, il fatto che entrino in gioco meccanismi di retroazione non deve ingannare in merito alla circostanza che esso, nell'ottica di Turing, rimane soltanto uno strumento inteso come utile integrazione del corpo umano. Arricchire le capacità senso-motorie non aggiunge nulla all'intelligenza, proprio come nulla aggiungerebbe un deltaplano che consentisse di volare[10].

La seconda ragione risiede nel fatto che il linguaggio, dal punto di vista di una spiegazione del *comportamento intelligente*, è indubitabile. È esso, infatti, che, in quanto concatenazione di simboli fonetici o grafici, mostra la produttività e la non ripetitività dell'intelletto umano senza dover affermare o dimostrare la propria esistenza, a differenza del pensiero, ostracizzato alla stregua di una chimera dalla riflessione della psicologia e della filosofia della psicologia nella prima metà del Novecento. Turing comincia a formulare le sue idee in merito alla macchina pensante in pieno clima comportamentista e l'influsso del comportamentismo è ben evidente nella formulazione del gioco dell'imitazione e della scelta del linguaggio come segno tangibile dell'intelligenza, il punto più elevato di una realtà neopositivisticamente configurata come gerarchia di livelli riducibili a quello fisico. D'altra parte, egli non nega l'esistenza del pensiero, contribuendo, come tutti gli scritti che in quel periodo appartengono agli albori dell'intelligenza artificiale, alla nascita del paradigma cognitivista e al rinnovamento dell'attenzione verso il pensiero come entità reale e realmente (scientificamente) indagabile. Infatti, la prima obiezione al gioco dell'imitazione che Turing considera, da lui stesso definita forte, è la seguente: «non possono forse le macchine comportarsi in qualche maniera che dovrebbe essere descritta come pensiero ma che è molto differente da quanto fa un uomo?» (Turing, 1950, p. 169). Quello che manca non è, dunque, una concessione di esistenza al pensiero[11], quanto piuttosto un riconoscimento del ruolo centrale dell'indagine sulla

[9] Le *condizioni* a cui il test può essere condotto coincidono con le *restrizioni* di base che il *modello di intelligenza* artificiale, cioè la macchina "pensante" nel senso di Turing, deve avere come *legittimo* partecipante al gioco dell'imitazione. La *praticità* che viene esclusa è qui da intendersi, verosimilmente, riferita a prestazioni senso-motorie, escluse per definizione dal gioco.

[10] Due precisazioni sono necessarie. Innanzitutto, dal punto di vista neuroscientifico è comprovato da più ricerche che l'attività senso-motoria del cervello influenza le altre attività cerebrali relative ad altre capacità mentali. La portata di questa influenza è ancora oggetto di indagine. In secondo luogo, ci possono essere modelli di IA che sfruttano meccanismi di retroazione più o meno complessi, a diversi livelli e in interazione fra loro. Anzi questa sembra essere una delle vie più promettenti nell'ambito di una modellistica fortemente simulativa dei fenomeni cognitivi.

[11] Forse può essere considerato un altro argomento contro il comportamentismo anche l'ultima della serie di obiezioni che Turing immagina rivolte contro la tesi della possibile individuazione di una macchina pensante attraverso il gioco

strutturazione dei processi di pensiero come linea guida per la simulazione al calcolatore dell'intelligenza, come testimoniato dalle parole con cui Turing commenta l'obiezione: «come minimo possiamo dire che se, ciononostante, una macchina può essere costruita in modo da giocare il gioco dell'imitazione soddisfacentemente, non abbiamo bisogno di tenerne conto» (*ibidem*)[12].

Ritornando al gioco dell'imitazione, va fatto notare come secondo Turing, per sua stessa ammissione, esso non rappresentasse un criterio ultimo e definitivo in merito alla presenza (o alla assenza) dell'intelligenza (del *pensiero intelligente*) in una macchina, come in seguito è stato inteso il Test di Turing. Il gioco dell'imitazione è appunto un *gioco*, che serve a mettere alla prova non una macchina, bensì un umano, il quale nel ruolo di interrogante deve riuscire a distinguere senza farsi ingannare il genere maschile o femminile del suo interlocutore oltre il muro, cioè all'altro capo dell'interfaccia di cui si serve per comunicare in forma linguistica. La percentuale di successo da parte dell'interrogante deve risultare significativamente vicina a quella che si verificherebbe se a giocare il gioco fossero un uomo e una donna e non una macchina e una donna. Inoltre, qualora ciò accadesse, vale a dire, qualora il gioco funzionasse e il test venisse superato, questo di per sé non costituirebbe «un criterio necessario per l'attribuzione di intelligenza, né, forse, sufficiente; non esiste neanche un modo chiaro per definire "superato" il test, ma solo la possibilità di stabilire "giocate" più o meno buone, nel senso di difficoltà di riconoscimento per l'interrogante più o meno paragonabili al caso di un interlocutore umano» (Lolli, 1994, p. 18).

Nel corso degli anni sono state proposte molteplici variazioni del Test di Turing, anche ironiche, come quella di Gunderson (1964) che è analoga a quello che succede nel gioco dello "schiaffo del soldato", in cui qualcuno viene colpito su una mano mentre è voltato di spalle e deve indovinare chi o che cosa lo ha colpito. Gunderson, infatti, propone di considerare come legittima la domanda: "possono pensare le pietre?" sulla scia del fatto che potrebbe essere molto difficile distinguere se il piede che abbiamo messo oltre una porta socchiusa, *fuori della stanza*, sia stato pestato da un uomo o colpito da una roccia che cade. Tralasciando le derive più banali di questa impostazione del Test di Turing, essa pone tuttavia l'accento sull'aspetto più comportamentistico della sua formulazione standard. Come fa notare Bara (1978) sono opportune alcune revisioni del Test perché esso funzioni e perché si possa definire con precisione cosa vuol dire averlo superato. La più importante è forse la versione estesa del Test proposta da Abelson: ETTA (*Extended Turing Test by Abelson*,)[13]. Per Bara, il passo fondamentale compiuto da Abelson consiste nell'aver dato «esplicitazione formale di

---

dell'imitazione, cioè quella relativa alla percezione extrasensoriale. Può lasciare sconcertati il fatto che Turing ne accetti l'esistenza, ma non si può eccepire sul fatto che, comunque si considerino i "poteri extramentali", accogliendoli si ammette per definizione l'esistenza di una mente.

[12] Un'opinione diversa esprimeranno in merito Newell e Simon in quella che può essere considerata un'embrionale formulazione del paradigma delle scienze cognitive. Requisito essenziale della simulazione al calcolatore delle attività cognitive era, a loro avviso, la riproduzione dei processi del pensiero umano e non solo il conseguimento di medesimi risultati. Per tale ragione la loro metodologia di ricerca consisteva nel collezionare resoconti di soggetti umani intenti a risolvere problemi al fine di ricavarne utili euristiche da implementare in un solutore generale di problemi, poi realizzato con il GPS (*General Problem Solver*). Si veda Newell, Simon (1972).

[13] Si veda in proposito Abelson (1968).

un principio epistemologico basilare, troppo facilmente dimenticato: il programma non deve riprodurre *tout court* un uomo, ma un suo modello» (Bara, 1978, p. 78). L'operazione compiuta da Abelson può essere considerata perciò una delle pietre miliari sulla via della modellizzazione cognitiva, che, non va dimenticato, si sviluppa come paradigma metodologico qualche anno dopo la nascita effettiva dell'IA.

D'altra parte, non si deve neanche dimenticare che il gioco dell'imitazione non fu concepito da Turing come Test. Lo divenne in seguito, dapprima come approdo teorico ultimo, come idea regolativa della ricerca, sulla via della realizzazione del comportamento intelligente da parte di una macchina, generalmente un calcolatore sul quale viene implementato un qualsivoglia tipo di programma o sistema di programmi; successivamente, come prova pratica da superare in una sfida fra diverse "macchine intelligenti"[14]. Ma anche se il gioco è stato considerato, dopo la morte del matematico britannico, un Test, e con questo nome si è conservato nella letteratura, «di tale termine Turing non fa mai uso» (Lolli, 1994, p. 17).

Ciò appare comprensibile se, ancora una volta, si guarda al periodo in cui Turing scriveva. Il maggiore interesse primigenio dei ricercatori pionieri nel campo dell'IA era rivolto ai giochi, per una serie di ragioni che vanno dalla ristrettezza, e quindi manovrabilità, del loro dominio all'impiego di strategie di ragionamento facilmente descrivibili da parte dei giocatori. Per un periodo di tempo relativamente esteso ancora prima che venisse coniato il termine "intelligenza artificiale"[15] nel periodo "preistorico" dell'IA, l'attenzione dedicata a giochi, quali il tic-tac-toe, meglio conosciuto in Italia come tris, la dama o i ben più filosoficamente connotati scacchi, fu enorme e pervasiva, anche grazie al libro di Von Neumann e Morgenstern dedicato alla teoria dei giochi[16]. Tuttavia, sebbene il fatto che Turing parli di *gioco* dell'imitazione è spiegabile in riferimento allo spirito che animava le prime ricerche in IA, si trattava pur sempre di una forma peculiare di gioco, un gioco *sui generis* appositamente creato, o perlomeno modificato, per saggiare le capacità di un programma. Questo, peraltro, non deve indurre a credere che fosse concepito come un test formale e diretto. Le capacità di cui un programma, che giocasse ragionevolmente bene il gioco dell'imitazione, potrebbe foggiarsi, non vanno intese come capacità cognitive in senso stretto, ma in un senso più generale di manifestazione complessiva di comportamento intelligente attraverso il linguaggio.

---

[14] È questo il famoso *Loebner Prize* che dal 1989 assegna medaglie ai programmi che si sono rivelati più intelligenti utilizzando come criterio di decisione il Test di Turing. Nessun programma fino ad oggi ha mai superato pienamente il Test e, di conseguenza, la medaglia d'oro non è mai stata assegnata. Sono state più volte conferite medaglie di minor pregio a riconoscimento della realizzazione di parziali abilità da parte di programmi. Sulla travagliata storia di questo Premio e per una parziale rassegna dei giudizi espressi in merito alla validità a fini scientifici di questo tipo di competizione si veda l'articolo di Sundman (2003) reperibile online al sito:
http://www.salon.com/tech/feature/2003/02/26/loebner_part_one/

[15] Nel famoso seminario di Dartmonth del 1956, in cui venne scelta questa dicitura a indicare una serie di ricerche che si differenziavano per metodi, impostazione e discipline di afferenza dei singoli ricercatori impegnati, ma che vertevano tutte sul comune obiettivo di ricreare prestazioni (simulazioni o emulazioni) intelligenti da parte delle macchine artificiali a quel tempo più avanzate, i calcolatori.

[16] Von Neumann, Morgenstern (1944). Sulla teoria dei giochi e il suo influsso sulla nascita dell'IA si rimanda a Franchi (2004).

Che cosa ci autorizzano ad affermare circa l'argomento della stanza cinese di Searle queste considerazioni in merito alle idee di Turing? Prima di dare una risposta a questo interrogativo è bene considerare un altro argomento che non menziona nessuna stanza, ma anticipa, condividendone l'impostazione, i presupposti teorici di quello di Searle.

1.4 Putnam e il telepate giapponese

In un saggio del 1975 dal titolo *Linguaggio e filosofia*[17] Hilary Putnam propone di immaginare una situazione di questo tipo. Si consideri un romanzo scritto in giapponese attraverso la tecnica narrativa del flusso di coscienza e un uomo che, privo di qualsiasi conoscenza della lingua giapponese, ne impari a memoria un brano piuttosto lungo. Si sottoponga, in seguito, questo individuo a una seduta di ipnosi in cui gli si comandi di ripetere mentalmente il brano appreso mnemonicamente «con tutte le giuste pause, intonazioni, enfasi, ecc. Se il suo comportamento non entra in aperto contrasto con quanto gli passa per la mente, in un certo senso sarebbe *come se* "pensasse in giapponese"» (Putnam, 1975, p. 25). Questo, secondo Putnam, è vero al punto che anche un telepate di madrelingua giapponese, potendo cogliere il flusso dei pensieri dell'individuo, lo scambierebbe per un individuo che pensa in giapponese. Tramite suggestione postipnotica, ci si potrebbe spingere fino a indurre l'uomo a credere di pensare in giapponese, così che anche le sue credenze in merito a ciò che sta facendo non potessero essere indizi rivelatori per il telepate del fatto che l'individuo non capisce affatto il giapponese. Nonostante questo, «è tuttavia chiaro che egli non penserebbe le proposizioni espresse dagli enunciati che gli attraversano la mente, dal momento che *in realtà* non *comprenderebbe* (quale che sia il suo "senso di comprensione") quegli enunciati» (*ibidem*).

La situazione appena descritta richiama quella che Searle immagina in *Menti, cervelli e programmi*. Infatti, non è irragionevole pensare che egli si sia ispirato a Putnam, il quale, a sua volta, definisce il racconto del telepate giapponese, un *Gedankenexperiment*. Ci sono, altresì, alcune differenze notevoli. In primo luogo, Putnam non si preoccupa delle implausibilità di cui arricchisce l'argomento, riscontrabili, ad esempio, nella effettiva possibilità di imparare a memoria un brano di una lingua che non si conosca affatto, incluse le intonazioni e le enfasi con cui il brano deve essere letto o, stando all'esperimento, *ripetuto mentalmente*. Si può superare uno scoglio di questo genere, però, dicendo che l'uomo che impara a memoria il brano del libro lo fa ascoltando i suoni pronunciati da qualcuno che possa capire e leggere il romanzo ad alta voce, quasi come se imparasse una canzone o una melodia o una generica successione di suoni. D'altra parte, anche ammesso che il flusso di coscienza in quanto tecnica narrativa sia in qualche maniera identico al flusso *della* coscienza, e di questo non si può dare che un'evidenza di tipo introspezionista, ciò che

[17] Il saggio è contenuto in Putnam (1975).

sembra ancor meno verosimile è l'utilizzo della telepatia come strumento in grado di cogliere tale flusso, consistente in una serie di enunciati nella mente dell'individuo totalmente ignorante del giapponese[18]. Appare abbastanza evidente che l'intento di Putnam è un altro, rispetto a quello di descrivere una situazione reale. Egli piuttosto avanza l'idea che, pur in una situazione palesemente assurda, non viene meno il fatto che la semplice enunciazione di alcuni enunciati, o il semplice pensarli nella mente[19], non bastano ad autorizzarne la comprensione da parte del parlante, o del pensante. Quest'ultima, piuttosto, deve essere vista «nel fatto che un parlante *che comprende* può *fare delle cose* con le parole e con gli enunciati che pronuncia (o che pensa nella propria testa), *oltre al* semplice pronunciarli» (*ibidem*).

Tutto ciò è ben diverso dallo scopo che si prefigge Searle con il suo argomento. Il suo obiettivo è piuttosto una critica nei confronti della tesi che afferma la possibilità di ricreare il pensiero intelligente attraverso l'implementazione di un programma che *generalmente* manipola simboli. Tale implementazione, ricordiamolo, dovrebbe dotare il computer su cui viene compiuta di stati cognitivi. Questi nell'ottica di Searle sono gli stati cognitivi che corrispondono *direttamente* alla fattiva possibilità del comprendere e la possibilità che si realizzino in questo modo è da lui esclusa. Tuttavia, il debito di Searle nei confronti del *Gedankenexperiment* di Putnam è decisamente esplicito. Consideriamo, perciò, da vicino le formulazioni dei due argomenti così che ciò in cui differiscono si renda evidente proprio attraverso l'analisi di ciò che li fa apparire simili.

La padronanza linguistica. In entrambi gli argomenti tutto ruota attorno alla (non) conoscenza di una lingua intuitivamente molto diversa dall'inglese, in un caso il giapponese, nell'altro il cinese. Si tratta in entrambi i casi di veicolare l'idea di una situazione palesemente controintuitiva, vale a dire la padronanza di una lingua sconosciuta, manifestamente difficile perché estremamente differente se raffrontata alle lingue occidentali e dotata di un diverso sistema di scrittura, per poi mostrare che tale padronanza è fittizia e si riduce a mero fatto esteriore, puramente meccanico, realizzabile attraverso un metodo. Nell'esperimento di Putnam, però, il metodo è mnemonico e nulla vieta che una persona possa attuarlo ad opportune condizioni, quali l'ascolto di una persona di madrelingua giapponese che legga con la giusta intonazione i brani del libro per un numero finito di volte, ma bastevoli a che l'individuo ignorante del giapponese possa apprendere la serie di enunciati che compongono il flusso di coscienza[20]. Nella situazione descritta da Searle, invece, non si tratta di

[18] Certo, a meno che non si voglia ammettere che Putnam, come già Turing, creda realmente nell'esistenza di poteri legati alla percezione extrasensoriale, circostanza quanto meno assai dubbia.

[19] Non va confusa, per ovvie ragioni, la ripetizione di enunciati nella mente, in una sorta di «monologo interiore», con il Linguaggio del Pensiero, il Mentalese, teorizzato da Jerry Fodor, il quale, plausibilmente, sfuggirebbe ai poteri del telepate, a meno che questi, essendone a conoscenza come tutti data la natura innata del Linguaggio del Pensiero, non sintonizzasse la sua "antenna telepatica" su questa "frequenza linguistica". Ma forse qui ci stiamo spingendo troppo oltre, facendo decadere l'accettabilità del *Gedankenexperiment*. La situazione che descrive Putnam va vista come analoga a quella in cui a volte ci ripetiamo interiormente filastrocche senza senso o testi di canzoni in una lingua che non conosciamo, ma che abbiamo imparato dopo ripetuti ascolti.

[20] Una difficoltà pratica, ma irrilevante ai fini teorici dell'esperimento, potrebbe consistere nella vaghezza con cui gli enunciati sono delimitati ai loro margini, e perciò sintatticamente ambigui, nella tecnica del flusso di coscienza.

ingannare un interrogante attraverso tecniche mnemoniche, la cui validità non può essere messa in dubbio, ma per mezzo di una corretta interazione in un rapporto di scambio reciproco in forma linguistica, che avviene sulla base di domande e risposte. Si può, di conseguenza, concludere che, per questo aspetto, Putnam e Searle mettano in campo *due tipi diversi di padronanza linguistica*.

La funzione dell'interrogante[21]. La differenza fra la situazione descritta da Putnam, di un telepate che "legge il pensiero", e quella descritta da Searle, di un individuo di madrelingua cinese fuori della stanza che legga le risposte fornite dal Searle chiuso nella stanza, è soltanto apparente e non deve trarre in inganno. In entrambi i casi siamo in presenza di *lettori di stringhe di simboli*, elementi costitutivi delle due lingue. Nel caso dell'esperimento della stanza cinese la cosa è evidente: si tratta di un individuo che riceve fogli scritti con simboli cinesi. Nel caso di Putnam la questione è più velata, ma occorre ammettere che non c'è alternativa a tale spiegazione. Infatti, il telepate nell'«ascoltare – per dirla con Putnam – il "monologo interiore"» dell'individuo che ha memorizzato il brano in giapponese corrisponde in tutto e per tutto all'interrogante di madrelingua cinese che si trova a leggere fogli pieni di simboli che rappresentano parole ed enunciati in cinese. Non si vede cos'altro potrebbe fare, se non leggere il pensiero, cioè i simboli del pensiero, se si vuole che l'argomento regga e sia utile al suo scopo, che è quello di dimostrare la pura esteriorità delle forme simboliche nella comprensione del linguaggio. In effetti, non fornendo, perché irrilevante, una spiegazione ulteriore e più approfondita della natura dei poteri telepatici dell'individuo di madrelingua giapponese, Putnam implicitamente invita ad assumere che la telepatia sia non altro che la percezione di una mera successione simbolica all'interno delle altre menti, senza alcuna potenzialità aggiuntiva[22]. La funzione dell'interrogante, o la parte essenziale della funzione dell'interrogante, consiste, dunque, nella comprensione, per via della sua natura di madrelingua, di stringhe di simboli della propria lingua.

La natura del metodo. Ciò che è problematico per Putnam, non lo è per Searle. Infatti, è parte essenziale del *Gedankenexperiment* della stanza cinese la presenza di un manuale di simboli della lingua sconosciuta uniti a una serie di istruzioni scritte nella propria lingua che permettano la correlazione dei simboli cinesi fra i tre plichi che vengono immessi nella stanza. Questo permette al

---

Tuttavia, si può supporre che questo problema sia superato nel momento stesso in cui l'individuo, che ignora il giapponese, decida di imparare a memoria non leggendo il testo, ma ascoltando le parole del lettore giapponese, il quale inevitabilmente dando intonazione al brano dà luogo in maniera implicita a una qualche disambiguazione sintattica del testo.

[21] Utilizzo il termine "interrogante" sia per il telepate di Putnam che per il tizio cinese o la squadra di programmatori che conosce il cinese fuori della stanza in cui è idealmente racchiuso Searle, anticipando il tal modo i termini per il confronto dei due argomenti con il gioco dell'imitazione di Turing.

[22] Questo non è certo una dimostrazione forte della descrizione/spiegazione del pensiero in termini di elaborazione di simboli. Se proprio si vuole è la congettura debole (perché necessariamente non supportata da "prove telepatiche") della presenza in una qualche parte della mente di un esatto corrispondente interiore del linguaggio esteriore, intendendo con "esatto corrispondente" una relazione biunivoca da simbolo a simbolo. Tale concatenazione simbolica interiore non è la stessa cosa che elaborazione, né, in forza di questo argomento, è detto che vi sia soggetta.

Searle rinchiuso di poter rispondere alle domande che gli vengono poste dall'esterno. In altri termini, questa serie di istruzioni, che Searle chiama "il programma", rende possibile l'*interazione* su base linguistica con l'esterno della stanza, il tutto in modo inconsapevole per quanto riguarda la lingua cinese da parte di chi agisce in base a quelle istruzioni. Nella situazione descritta da Putnam non c'è nulla di simile. Non esiste, cioè, un *metodo* formalizzato per l'interazione. Al massimo si può pensare a un metodo mnemotecnico per apprendere il brano in lingua giapponese, come si è voluto suggerire ipotizzando l'idea di un lettore giapponese, l'ascolto del quale permetta all'individuo che non conosce la lingua di memorizzare il brano scelto casualmente. E non c'è un metodo che permetta l'interazione per la semplice ragione che non c'è interazione. Il problema di captare il monologo interiore ricade tutto nelle possibilità e nelle capacità del telepate. Fra i due non esiste un vero scambio linguistico. Se ci fosse, dimostrerebbe proprio ciò che Putnam nega sia possibile attraverso il metodo del mandare a memoria, cioè il *saper fare qualcosa con le parole* al di là della loro mera enunciazione per imitazione. Per Searle, invece, non è problematico il fatto che sia possibile dotare l'individuo nella stanza di un programma composto da una serie di istruzioni per rendere attuabile l'interazione con l'esterno. In altri termini, non viene problematizzata la *costruzione del metodo in cui viene reso possibile lo scambio in forma linguistica di domande e risposte*, che, al contrario, costituisce uno degli obiettivi dell'IA.

Se ne può concludere che, se un individuo chiuso in una stanza e in una situazione come quella descritta da Searle è verosimilmente inconsapevole delle operazioni che sta compiendo e, quindi, non ha comprensione alcuna dei simboli linguistici a lui sconosciuti che sta manipolando, non è così ovvio come possa essere costruito l'insieme delle istruzioni che rendano, invece, plausibile dall'esterno una reale interazione con l'uomo nella stanza. Il problema non sta in chi manipola le istruzioni, ma in chi le formula, cioè in chi progetta, organizza e costruisce il metodo. Il fatto che Searle lasci in ombra tale questione di difficile risoluzione indica il tentativo da parte sua di rendere plausibile ciò che plausibile non è, al contrario di Putnam che non nasconde gli aspetti irrealistici del suo esperimento mentale (la telepatia) proprio perché irrilevanti ai fini di ciò che intende sostenere. La stessa spiegazione della comprensione è diversa nei due filosofi. Se per Putnam essa risiede nella possibilità di attuare determinate pratiche attraverso il linguaggio (concezione che può essere ricondotta, con la dovuta cautela, a quella del "significato come uso"), per Searle alla negazione della possibilità di comprensione da parte di un programma corrisponde l'assunzione aprioristica che la comprensione del linguaggio è qualcosa che un individuo attua grazie ai propri non specificati poteri causali del cervello.

In conclusione, l'esperimento della stanza cinese di Searle può essere facilmente ricondotto al *Gedankenexperiment* di Putnam, in base a manifeste analogie e a un superficiale omeomorfismo di costruzione. Le differenze di fondo che si sono evidenziate dovrebbero aver chiarito una maggiore inattaccabilità del secondo di contro a una debolezza intrinseca del primo, fatta scivolare in secondo

piano attraverso l'artificio della plausibilità intuitiva dei passaggi fondamentali di cui l'argomento si costituisce. Fin qui si è mostrato che quello di Putnam può essere considerato, dal punto di vista della sua struttura, come un antecedente più o meno implicito dell'argomento di Searle. In che modo c'entra Turing?

1.5 Lo spostamento della prospettiva

Nello scritto di Searle il riferimento al saggio di Turing, *Macchine calcolatrici e intelligenza*, è indiretto, ma non completamente celat,. Se ne ritrova traccia, in particolare, nelle intenzioni che compongono i suoi obiettivi, nella scelta del tema, nell'andamento dell'argomentazione. In precedenza, abbiamo già accennato all'importanza che il linguaggio riveste nel gioco dell'imitazione di Turing ed è anche manifesta la centralità del suo ruolo nel *Gedankenexperiment* della stanza cinese. Tuttavia, questo accostamento va indagato ulteriormente e analizzato nei punti di contatto, affinché, ancora una volta, ne risaltino le incongruenze.

Occorre dire, innanzitutto, che l'obiettivo polemico di Searle, come già ricordato, è il gruppo di ricerca di Yale guidato da Roger Schank, così come i programmi che in qualche modo sono costruiti con l'obiettivo di simulare la comprensione del linguaggio naturale, quali, ad esempio, ELIZA di Weizenbaum e SHRDLU di Winograd[23]. D'altra parte, egli afferma che «i suoi argomenti si applicherebbero, [...] in pratica, a qualunque simulazione da parte di una macchina di Turing dei fenomeni mentali umani» (Searle, 1980, p. 47). Si tratta, perciò, di una tesi contro il computazionalismo classico come spiegazione dei processi di pensiero, che Searle chiama "intelligenza artificiale forte". Non c'è spazio, d'altra parte, per un'intelligenza artificiale debole nella concezione di Searle, il quale considera quale radice unica di tutti i «fenomeni mentali umani» i non meglio precisati «poteri causali del cervello», a meno di intendere questo secondo tipo di IA come un ridimensionamento degli obiettivi più che delle pratiche – delle strutture algoritmiche, delle teorie computazionali, dei modelli simulativi – dell'IA.

In ogni caso, un attacco diretto a Turing non è presente in maniera esplicita nello scritto di Searle. Il generalizzare la sua critica a «qualunque simulazione di una macchina di Turing», se può essere fatto valere come una critica al Test di Turing, lo è in ragione della centralità attribuita nel suo argomento al linguaggio. Eppure nel gioco dell'imitazione di Turing il linguaggio è solo il medium espressivo-comunicativo fra l'interrogante e la macchina (che si finge uomo), la quale esibisce capacità linguistiche in quanto *segno esteriore* di tutte le attività cognitive. In altri termini, Turing non ci dice nulla a proposito del modo in cui una macchina possa produrre il linguaggio coma capacità cognitiva in aggiunta alle altre – memoria, ragionamento deduttivo o induttivo, formulazione di ipotesi, costruzione di analogie, astrazione e creatività in ambiti diversi quali la

[23] Si vedano Weizenbaum (1965, 1978) e Winograd (1972, 1973).

matematica o la poesia – che dovrebbe esibire in una conversazione con un interrogante umano; né ci informa sul ruolo occupato in un sistema cognitivo dalla prestazione linguistica; né afferma, infine, alcunché in merito agli antecedenti psicologici o alle radici logiche del linguaggio. L'obiezione di Searle alla possibilità del computazionalismo come teoria esplicativa dei processi mentali è, dunque, basata sull'attribuzione a Turing di un intento superiore a quello che quest'ultimo si prefiggeva con il gioco dell'imitazione, «che non si riferisce a singole capacità, che non richiede un esperto come interrogante, che non propone una prova da superare da parte della macchina, ma una prova da superare da parte degli interroganti rispetto alle macchine» (Lolli, 1994, p. 18). Inoltre, l'eventuale superamento della prova non avrebbe di certo giustificato la presenza di una qualche particolare attività cognitiva nella macchina. Questo perché, diversamente rispetto al *Test*, che ruota attorno alla macchina, il *gioco* dell'imitazione ruota attorno all'uomo (l'interrogante) e può essere considerato, più che un test per verificare l'intelligenza delle macchine, un *Gedankenexperiment* per vagliare l'atteggiamento umano di fronte alla simulazione delle prestazioni intelligenti, il che equivale a dire, una valutazione dei principi teorico-epistemologici alla base dell'impresa dell'IA, sia essa simulativa o emulativa, simbolica o connessionista, rappresentazionalista o dinamica o situata o di qualsiasi altro tipo.

Naturalmente, questo non fa di Turing un teorico del connessionismo. Nel saggio *Intelligent Machinery*[24], egli si era già dimostrato tutt'altro che disinteressato sia alla questione dell'apprendimento automatico sia al problema del cervello e della natura *continua* del suo funzionamento, di contro alla natura *discreta* dei calcolatori e in generale di tutti gli automi a stati finiti. Ma Turing scrive nell'epoca della nascita dei calcolatori e non può non essere colpito dalle enormi possibilità che si aprono grazie al loro sviluppo e al loro impiego perfino in un ambito di studi come quello delle scienze della mente. Il suo pensiero era senz'altro più aperto di quello dei suoi prosecutori rispetto alle contrapposizioni, talvolta meramente di stampo ideologico, che hanno caratterizzato il primo cinquantennio di storia dell'IA[25].

Il gioco dell'imitazione, perciò, non è, e non può essere, come testimoniano le intenzioni del suo ideatore, una sorta di *experimentum crucis* da predisporre ogniqualvolta si voglia mettere alla prova un programma in merito all'effettiva riproduzione o meno di una determinata capacità cognitiva. In quale maniera, dunque, si è arrivati a considerarlo tale? Quale operazione compie Searle, in riferimento ad esso, nella costruzione del suo argomento?

L'operazione che porta dal gioco dell'imitazione alla stanza cinese consiste di due passi fondamentali. Il primo è la trasformazione del gioco dell'imitazione nel Test di Turing, la quale costituisce uno spostamento di prospettiva all'interno dell'IA. Metaforicamente, esso può essere

[24] Turing (1948).

[25] Contrapposizioni di coppie di concetti dualistici che affondano le loro radici in remote dispute filosofiche e che sembrano lontane dall'essere risolte, così come l'IA sembra ancora lontana da un'emancipazione completa dai dualismi concettuali che di volta in volta le fanno trascurare alcuni aspetti a scapito di altri, più che considerarli di pari importanza e affrontarli nel modo migliore e più proficuo.

considerato come un'uscita dalla stanza da parte dell'interrogante con la conseguente sostituzione al suo posto del modello cognitivo (il programma), che nel gioco ricopre la funzione di "entità" la cui natura va indovinata. Questo scambio di posti non è di poco valore, perché ad esso corrisponde, *dal punto di vista della giustificazione teorica, l'inversione dell'onere della prova*, non più a carico dell'uomo, ma a carico del programma. Di conseguenza il gioco perde la sua natura di gioco – non è più un interrogante umano a dover indovinare se sta dialogando con un uomo oppure con una macchina – per diventare esperimento cruciale in cui il programma deve dimostrare di possedere e mettere in pratica una o più capacità cognitive.

Tale cambiamento di prospettiva espone l'IA, un'IA che si avvalga del Test di Turing come del suo esperimento cruciale, ad accuse di operazionalismo e, ancor di più, di comportamentismo, che Searle non manca di sottolineare (Searle, 1980, p. 69-70). Tuttavia, questo non significa che l'IA non sia riuscita a staccarsi da una visione comportamentistica del mentale per quanto riguarda l'analisi dei risultati ottenuti. Al contrario, ciò sarebbe equivalso ad una sorta di eliminazionismo, il quale non costituisce di certo uno degli indirizzi prevalenti dell'IA sia simbolica, uno degli obiettivi principali della quale è l'indagine del meccanismi del *pensiero*, sia connessionista, che pure non può essere posta del tutto al di fuori della cerchia del funzionalismo[26] con tutto il vocabolario teleologico e mentalistico che esso implica per le spiegazioni fornite dalle scienze cognitive. Tali accuse, invece, evidenziano il fatto che il Test di Turing non è, e non può essere, uno strumento completo ed esaustivo di valutazione in merito al raggiungimento di un obiettivo prefissato attraverso la costruzione di un modello cognitivo[27], allo stesso modo in cui il semplice conversare con qualcuno non ci svela, né può farlo, la natura dei meccanismi alla base delle sue capacità cognitive, consentendoci al massimo la mera attribuzione, mai del tutto assoluta, dell'effettiva presenza nel nostro interlocutore della capacità di assolvere ad alcune prestazioni (intelligenti). Per assurdo, un programma che superi il Test di Turing, e che non venga valutato come modello secondo altri parametri – quali ad esempio l'esame della struttura della sua architettura, della funzione delle sue componenti, del fine per cui viene progettato e delle restrizioni predefinite cui viene assoggettato – ha così poca possibilità di dirci qualcosa sulla natura del pensiero umano, quanta ne ha in misura inversa di essere accettato in una società di individui umani, essendo il linguaggio il più potente mezzo di interazione e di socializzazione fra individui. Ma non è sufficiente la natura non privata e sociale del linguaggio a descrivere tutti i processi mentali, così come appare altrettanto lontana dal riuscire nell'intento di una loro descrizione utilizzare un punto di vista esclusivamente neurofisiologico.

Se la critica di Searle si arrestasse a questo, non sarebbe del tutto fuori luogo, mettendo in guardia l'IA dal rischio di confondere la duplicazione o l'emulazione del comportamento con la spiegazione del fenomeno duplicato. Egli, però, si spinge oltre. Il secondo passo dell'operazione di

[26] Per una valutazione del ruolo del funzionalismo nelle scienze cognitive si veda Cordeschi (2002).
[27] Si veda quanto già detto in proposito nel paragrafo 3.

trasformazione del gioco dell'imitazione può essere visto come un nuovo ribaltamento del punto di osservazione. Mentre in un primo momento c'era stato un metaforico scambio di posti fra interrogante e macchina, ora c'è un ritorno indietro dell'interrogante – immaginiamo che sia Searle stesso – il quale rientra nella stanza dove precedentemente erano stati messi alla prova, *giudicati*, prima l'interrogante e poi la macchina pensante. Ma il suo rientrare è simultaneamente un entrare nella macchina, che si trova ancora nella stanza. Da questo nuovo punto di vista, Searle può svincolarsi dal ruolo che prima era tipico dell'interrogante, trasformandone la funzione, e giudicare la macchina dal suo interno. Si hanno, in tal modo, due tipi di interrogante nell'argomento della stanza cinese: 1) un interrogante di primo livello, impersonato dalla figura del madrelingua cinese o dagli individui che compongono il team di programmatori-interroganti in lingua cinese, i quali, dall'esterno, *non possono comprendere la mancata comprensione* del cinese da parte del Searle manipolatore all'interno della stanza; 2) un interrogante di secondo livello, il Searle chiuso nella macchina, che ne osserva il funzionamento da dentro, anzi che diventa parte dello stesso funzionamento, e che afferma di non comprendere nulla di quello che sta facendo, se non che sta compiendo operazioni formali su simboli a lui ignoti.

Vediamo questo a cosa conduce. Mentre il ruolo dell'interrogante di primo livello corrisponde a quello del telepate giapponese nel *Gedankenexperiment* di Putnam, il secondo livello di interrogazione è la mossa decisiva che Searle muove nei confronti della tesi di Turing sulla possibile esistenza di macchine pensanti. Una macchina non può pensare, o esibire capacità cognitive, perché una macchina non può arrivare a comprendere quello che sta facendo, e in questa situazione specifica non può arrivar a comprendere i *concetti* del linguaggio che sta producendo. Infatti, è intuitivo che non si dia effettiva comprensione dei segni che si stanno manipolando, se la loro manipolazione avviene attraverso regole esplicite la cui applicazione può essere attuata per mezzo di un mero raffronto di forme figurative (le forme dei segni sui fogli che compongono la storia con quelle dei segni sul manuale). Ciò che rimane inesplicata, invece, è la natura delle regole che compongono il manuale di istruzioni, le quali, si è già detto, costituiscono il vero problema, il cui superamento può assurgere a emblema di ogni obiettivo di fondo dell'IA.

Il fatto che qualcuno possa osservare la realtà interna di un meccanismo, capirne il funzionamento e, tuttavia, essere estraneo alla comprensione del fenomeno prodotto, non è un'obiezione stringente in senso assoluto nei confronti delle possibilità dell'IA, anche se la sua attività all'interno della stanza costituisce una parte essenziale e ineliminabile di tutto il processo. Piuttosto, tutto l'argomento può essere considerato come un *caveat* nei confronti della costruzione di modelli simulativi. Non si può, infatti, non tenere conto, nella costruzione della loro architettura, sia della scopo e della funzione delle singole parti che li compongono, sia delle restrizioni che un modello deve avere, da una parte, per essere l'effettiva simulazione di un processo, dall'altra, per evitare di diventare mera copia riproduttiva dell'originale. L'argomento della stanza cinese suggerisce che è sempre possibile trovare un livello di descrizione di un meccanismo totalmente al

di là, o, meglio, al di qua, della effettiva spiegazione dei processi posti in atto dal meccanismo. Infatti, l'argomentazione di Searle è mancante non tanto nel dimostrare che la mera manipolazione formale (interpretando tale termine nell'unica maniera sensata, ovvero nel senso di un formalismo logico-sintattico) non può portare alla comprensione del linguaggio e della sua natura di concatenazione di enunciati sintatticamente *e semanticamente* ben formati, quanto piuttosto nella lacuna relativa alla natura del manuale di istruzioni, "il programma", usato dall'individuo-Searle nella stanza. Che la comprensione del linguaggio naturale possa essere ridotta a un insieme di regole, espresse o meno in un linguaggio formalizzato, ma pur sempre regole, non sembra così plausibile come egli vuol far sembrare, anzi è un fatto piuttosto problematico. E non si vede come non si possa parlare di comprensione linguistico-concettuale all'interno di un'interazione comunicativa fatta di domande e risposte[28]. Searle non ci dice come sia possibile tale esplicitazione in regole formali, ovvero che agiscono esclusivamente in base alla forma dei simboli cui si applicano. Se ce lo dicesse, il suo argomento sarebbe invalidato[29]. Non dicendolo, lo espone a un forte rischio di implausibilità.

1.6 Le obiezioni alla stanza

Una volta costruito, l'argomento della stanza si presta a una serie di obiezioni, le quali, sia nel caso di Turing che in quello di Searle, sono state portate in prima battuta dagli stessi autori del rispettivo *Gedankenexperiment*.

Cominciamo da Turing. Egli immagina sia possibile portare all'idea di una macchina pensante, intendendo con questa accezione una macchina in grado di giocare al gioco dell'imitazione, una serie di obiezioni, quali:

a) l'affermazione che «il pensare sia [esclusivamente] una funzione dell'anima immortale dell'uomo» (Turing, 1950, p. 176), chiamata "obiezione teologica";
b) l'affermazione che «le conseguenze delle macchine pensanti sarebbero terribili» per l'umanità e, perciò, si spera in una loro irrealizzabilità (Turing, 1950, p. 177), definita "obiezione della 'testa nella sabbia'";
c) l'ipotesi che dimostrazioni logico-matematiche come quella del teorema di Gödel o ipotesi nell'ambito della matematica, ad esempio la tesi di Church-Turing, mostrino le «limitazioni

[28] A meno che, ancora una volta, le domande siano finalizzate all'applicazione di una serie di regole formali per ottenere una risposta, quali possono essere, ad esempio, domande relative all'applicazione di una qualche funzione su insiemi di numeri, come le operazioni del calcolo elementare. Tuttavia, anche questo modo di "seguire una regola" non è necessariamente univoco e può risultare estremamente diverso nell'uomo e nel calcolatore.

[29] Non si comprenderebbe più, fra le altre cose, la necessità esclusiva di una macchina con poteri speciali, come Searle definisce il cervello, affinché possa darsi la comprensione linguistico-concettuale. Basterebbe un calcolatore a un livello di complessità sufficiente per poter implementare linguaggi logico-formali del primo ordine.

[intrinseche] ai poteri delle macchine a stati discreti» (Turing, 1950, p. 178), che porta il nome di “obiezione matematica”;

d) l’affermazione che è possibile arrivare a sapere che una macchina pensa soltanto con l’essere quella macchina stessa e col «sentire se stessi pensare» (Turing, 1950, p. 179), convinzione sottoposta al giogo del rischio solipsistico e che Turing chiama “argomento dell’autocoscienza”;
e) l’opinione secondo cui se una macchina può fare qualcosa, allora quel qualcosa è, per definizione, “meccanizzabile” e per tale ragione privo di interesse, poiché non coglie il nocciolo reale del pensare, bensì solo alcune sue manifestazioni esteriori. Questa affermazione, che pecca di essenzialismo e rende asintotica la ricerca sui processi del pensiero, è in qualche modo analoga ad a) e d). Turing definisce genericamente questo modo di affrontare la questione “argomentazioni fondate su incapacità varie”;
f) la pretesa che le macchine possano fare solo ciò per cui sono programmate, cioè l’“obiezione di Lady Lovelace” nei confronti della macchina analitica di Babbage;
g) l’affermazione di una differenza incommensurabile fra la continuità del cervello e la natura a stati discreti dei calcolatori, che Turing battezza come “argomentazione fondata sulla continuità del sistema nervoso”;
h) l’affermazione che l’agire umano non è governato in tutti i possibili casi da regole fisse e prestabilite, come accade invece nel caso della macchine. Turing la definisce “argomentazione del comportamento senza regole rigide”;
i) la sorprendente idea che la presenza di un individuo dotato di poteri mentali particolari invalidi la possibilità di una corretta conduzione del gioco dell’imitazione, ovvero la già ricordata “argomentazione fondata sulla percezione extrasensoriale”.

È stato fatto notare che l’introduzione del gioco dell’imitazione da parte di Turing ha come fine primario quello di «discutere le obiezioni alla possibilità di costruire macchine pensanti» (Lolli, 1994, p. 19), piuttosto che quello di criterio di decisione in merito alla realizzazione effettiva di una macchina pensante, in seguito attribuitogli dalla letteratura con il nome di Test di Turing,. Le obiezioni, però, non sono tutte uguali. In a), b) ed e) troviamo espressi una serie di pregiudizi nei confronti delle macchine, non argomentati, né argomentabili, che hanno la forma del convincimento dogmatico; d) è una tesi filosofica, cui soggiace un soggettivismo estremo, e che, se portata alle sue estreme conseguenze, procurerebbe nell’ambito delle scienze cognitive un’impossibilità metodologica effettiva nei confronti di qualsiasi tentativo di indagine del mentale al di fuori dell’analisi introspettiva; i) è un *caveat* allo svolgimento del gioco, che pur nell’assurdità della sua formulazione, e del tutto indipendentemente da quello che ne pensasse Turing, permette di circoscrivere l’effettivo campo d’azione del gioco.

Le restanti quattro obiezioni sono di natura diversa e portano un attacco dall'interno al computazionalismo e all'IA in generale, ponendosi in qualche modo sullo stesso piano. Sono argomentazioni costruite a partire da dati di fatto, laddove le altre hanno un carattere squisitamente aprioristico[30]. In particolare, c), f) e h), sono tre sfaccettature di un'unica obiezione, quella che riguarda le limitazioni di ogni sistema logico-deduttivo basato su regole esplicite: la sua incompletezza a comprendere tutti gli aspetti della realtà. Tale incompletezza si esprime sia nell'insufficienza del sistema formale a poter produrre tutte le verità in esso stesso esprimibili, sia nell'incapacità di valicare la rigida sequenzialità e monotònicità dell'applicazione delle regole ai suoi enunciati (assiomi e teoremi). Tuttavia, proprio negli anni in cui Turing scriveva, la nascita dell'IA era il primo tentativo di superamento di tale monotònicità, da una parte attraverso lo sfruttamento delle possibilità conferite dai costituenti strutturali degli algoritmi, come la chiamata di procedura, la funzione di scelta condizionata e la ripetizione, dall'altra attraverso l'adozione del metodo euristico di ricerca nello spazio problemico[31]. Resta da considerare g), che denota la lungimiranza con cui Turing enuclea il problema matematico alla base della contrapposizione fra IA simbolica e IA connessionista, un problema che non tocca da vicino chi gioca il gioco dell'imitazione, ma pone in primo piano la questione delle restrizioni che devono essere tenute in conto nell'ideazione di un modello cognitivo, questione equivalente a quella dell'appropriato livello di descrizione del fenomeno da simulare.

È possibile che Turing non facesse distinzione di sorta fra le obiezioni elencate nel suo saggio, poiché in esso l'indagine sulla effettiva possibilità di una macchina pensante non è separata da quella relativa a quali condizioni è necessario fissate per poter fare un'affermazione del genere, quali pregiudizi devono essere superati, quali principi teorici costituiscono un avvertimento costante alla ricerca in IA senza che possano mai essere rigettati come semplici problemi passibili di una soluzione definitiva. Tuttavia, un ruolo centrale spetta alle quattro obiezioni "interne", c), f), g) e h), le quali, sia detto per inciso, sarebbero valide anche nel caso in cui un qualche programma superasse il Test di Turing.

Torniamo a Searle. Abbiamo descritto l'argomento della stanza cinese in quanto obiezione indiretta all'idea di Turing di una macchina pensante e abbiamo visto come esso possa essere considerato tale in due modi. Per un verso esso si configura come critica al carattere operazionalista e comportamentista del Test di Turing, accusa che sembra giustificata se riferita a una versione "superficiale e troppo fiduciosa" dell'IA, ma che trascende le reali intenzioni di Turing. Da un

[30] È necessaria una precisazione. L'"obiezione dell'autocoscienza", se interpretata in chiave non solipsistica, pone all'attenzione della modellizzazione cognitiva il problema della soggettività e della natura qualitativa dei fenomeni mentali. Se questi, i così detti qualia, non diventano baluardo dell'oltranzismo negazionista dell'IA, costituiscono un ottimo stimolo alla riflessione epistemologica sui principi dell'intera ricerca in questo campo.

[31] Teoria algoritmica e metodo euristico devono aver contribuito non poco alla diffusione dell'idea della possibilità di una macchina pensante, non solo dal punto di vista teorico e astratto, il punto di vista della Macchina di Turing (MdT), ma anche per quanto riguarda gli aspetti applicativi, cioè la realizzazione fisica di strumenti (*hardware*) sempre più potenti e in grado di dare un supporto alle macchine astratte, rendendo così possibile l'implementazione dei algoritmi che implicano un numero sempre più elevato di risorse di elaborazione.

punto di vista più diretto, l'argomento di Searle è un'obiezione contro la possibilità che un programma comunichi attraverso il linguaggio naturale e nel farlo metta in atto processi simili a quelli di un essere umano. La manipolazione formale di simboli esclude un processo di comprensione, dimostrando in tal modo l'effettiva non coincidenza dei processi del pensiero umano, legati ai poteri causali del cervello, con i procedimenti algoritmico-formali realizzati in un calcolatore. In questa seconda accezione, il collegamento con Turing è riscontrabile nel fatto che il gioco dell'imitazione è basato sull'utilizzo del linguaggio naturale. La mossa implausibile attraverso cui questa seconda tesi è costruita è stata sottolineata in precedenza. Rimangono ora da esaminare le obiezioni che Searle, come Turing, individua nei confronti del suo stesso *Gedankenexperiment*. Egli le suddivide in sei repliche possibili e le enuncia unitamente alla loro confutazione. Elenchiamole:

1) non è l'individuo che comprende il cinese, ma il sistema di cui l'individuo è soltanto parte. In definitiva, il sistema si riduce, però, a due soli elementi necessari, l'individuo e il manuale di istruzioni, "il programma", che possono diventare uno soltanto se l'individuo nella stanza interiorizza il "programma" memorizzandolo. Tale operazione non gli permette ancora di comprendere il cinese, bensì solo di *imparare a memoria un metodo*. Questa è la "replica del sistema";
2) il problema si risolve se prendiamo un robot che incorpora un calcolatore. Tuttavia, questo, pur potendo interagire con l'ambiente, non ha comunque stati intenzionali; in altri termini, immettere la stanza all'interno di un sistema senso-motorio in grado di avere percezioni e di compiere movimenti non dota il programma della capacità di comprensione. Il Searle nella stanza può continuare indisturbato le sue funzioni. Questa va sotto il nome di "replica del robot";
3) la soluzione sta nel progettare una macchina che simula tutte le sequenze di propagazione dell'attività neuronale del cervello di un cinese mentre parla cinese. Questa, però, non avrebbe ancora stati intenzionali. Si potrebbe immaginare, infatti, di sostituire il cervello con un sistema di tubature e valvole in cui scorre acqua, azionato da un Searle idraulico. Costui guardando il sistema non avrà la benché minima comprensione del cinese, esibito esternamente in forma linguistica dall'intero sistema, perché il sistema simula soltanto le proprietà "formali" neurobiologiche, non sufficienti a produrre quelle causali. È la "replica del simulatore del cervello";
4) le tre obiezioni precedenti, che falliscono singolarmente, acquistano forza se prese tutte insieme. Tuttavia, l'idea di un robot con un cervello simulato al suo interno al posto del calcolatore e considerato come un sistema complessivo sarebbe esposta alla stessa obiezione di 3): un uomo potrebbe celarsi nella stanza del cervello simulato (o, perché no, controllarlo

da lontano con un telecomando secondo apposite istruzioni). Questa è "la replica combinata";

5) la conoscenza che si ha della comprensione che gli altri hanno del cinese o di altre cose deriva dall'osservazione del loro comportamento. Lo stesso tipo di conoscenza si deve applicare ai computer se esibiscono lo stesso comportamento. Questa "replica delle altre menti" non è altro che un ritorno al Test di Turing, quindi suona come una *petitio principi*, o, almeno, come una confusione fra *demonstrans* e *demonstrandum*;
6) è possibile tralasciare l'impostazione computazionale e adottare una strategia diversa, sempre nell'ambito dell'IA, per riprodurre i procedimenti causali specifici del cervello. Questa viene definita come "replica delle molte sedi" e ha il difetto di non colpire nel segno, perché l'argomento della stanza cinese si applica solo alla versione computazionale («forte») dell'IA.

Si vede bene come queste sei obiezioni non sono tutte sullo stesso piano. La 5) e la 6) vengono rigettate come non dirette all'argomento. Tuttavia, con la 6) Searle sembra concedere una qualche possibile speranza all'IA non simbolica in senso classico, in tutte le accezioni possibili. Ma è una debole speranza. Infatti, la 3) è una presa di posizione contro la simulazione dei meccanismi cerebrali, e quindi verosimilmente contro il connessionismo[32], che trova appoggio nell'estensione del potere confutatorio della stanza cinese ad una supposta ma non ancora realizzata formalizzazione (vale a dire, traduzione in simboli e regole esplicite) di tutta l'attività neuronale. Non sembra interessante la 4) perché nulla aggiunge alle tre precedenti, non resistendo in tal modo agli argomenti con cui queste vengono rigettate. È interessante, invece, la 2), poiché con essa Searle esclude che la percezione e l'interazione con l'ambiente siano di una qualche rilevanza ai fini del verificarsi di stati intenzionali e della comprensione, una tesi che sembra accettabile solo entro certi limiti. In ogni caso, la 2) e la 3) sono solo estensioni della 1). Nella 2) la stanza è immessa in un robot al posto del calcolatore che lo controlla; nella 3) la stanza è l'interno del calcolatore che riproduce fedelmente i collegamenti sinaptici di un cervello che capisce il cinese.

Questo porta a considerare come obiezione originale soltanto la 1). Essa viene curiosamente rigettata da Searle con un procedimento che ricorda ancora il *Gedankenexperiment* di Putnam, la memorizzazione del manuale di istruzioni. Questo procedimento annullerebbe la presenza di un sistema complessivo costituito dal "Searle nella stanza" più "il manuale", cioè "il programma", e si darebbe il caso di un individuo con due sottosistemi, uno che gli permetta di comprendere l'inglese e un altro, all'interno del primo, che gli permetta di agire, di fare qualcosa, con i simboli cinesi. Il fatto che il secondo sottosistema è soltanto una parte del primo sta a significare che condizione necessaria e sufficiente per la memorizzazione delle regole (istruzioni) e dei simboli formali (le

[32] Occorre notare che al tempo in cui scriveva Searle poca attenzione veniva ancora riservata ai modelli simulativi basati sull'utilizzo delle reti neurali, che di lì a poco sarebbero diventati l'approccio predominante nella ricerca in IA e nelle scienze cognitive.

raffigurazioni grafiche degli ideogrammi cinesi) è la comprensione dell'inglese. La memorizzazione sarebbe solo un fatto esteriore, come lo era per l'individuo che mandava a memoria brani in giapponese scritti con la tecnica del flusso di coscienza. C'è, però, un doppio ostacolo. Anche ammettendo che tale procedimento di memorizzazione sia possibile con lunghi sacrifici (I ostacolo), l'implausibilità di tutto questo risiede ancora nella mancata verosimiglianza del manuale di istruzioni (II ostacolo), come si è fatto rilevare più sopra.

Searle non sembra difendere in maniera convincente il suo *Gedankenexperiment* dalle obiezioni che egli stesso avanza, anzi dalla obiezione 1), di cui le altre, a meno di non deviare dall'argomentazione principale, sono casi particolari. Nell'avanzare la sua tesi egli mostra di avere una teoria del mentale non giustificata, essendo i fenomeni mentali riconducibili ai poteri causali del cervello, i quale rimangono inesplicati. Allo stesso tempo, mostra di avere una eccessiva fiducia nella possibilità di ridurre il linguaggio a regole esplicite in base alle quali sostenere in maniera formale una conversazione fatta di domande la cui risposta deve essere per forza univoca e non ambigua. In caso contrario, l'ambiguità risalirebbe fino alle regole stesse[33]. Nonostante questo, il bersaglio di Searle è il computazionalismo, inteso come manipolazione formale di simboli, nell'ipotesi in cui esso venga considerato un'adeguata teoria del mentale. In base a queste tesi, e stando a quello che Searle afferma con l'argomento della stanza cinese, appare inevitabile che si debba procedere a un'esclusione del linguaggio naturale, per via della sua "semplice e immediata" riducibilità a regole esplicite, dall'insieme dei fenomeni mentali rilevanti. Questo, però, è proprio l'opposto di quello che Seale vuole ottenere con la stanza cinese.

Rimane, comunque, la sensazione che in qualche modo il suo argomento non debba essere rigettato per intero, ma abbia una qualche utilità. Esso, infatti, invita a porci alcune significative domande: che tipo di computazionalismo può essere sensatamente proposto come spiegazione dei processi mentali, visto che a un qualche livello esso deve necessariamente essere ammesso? Sulla base di quali assunti teorici è costruibile un'adeguata nozione di computazionalismo? Se esso è manipolazione, o elaborazione, formale di simboli, quale livello o quali livelli è opportuno indagare attraverso questa nozione teorica? Il cervello[34] o la mente? O entrambi? O qualcosa di intermedio?

Numerose sono state le reazioni immediate alla presentazione del saggio di Searle[35] e non ci interessa in questa sede una loro disamina completa. Prenderemo in considerazione, come ultima obiezione all'argomento della stanza, la critica che Hofstadter rivolge all'articolo di Searle a un

[33] C'è un'altra possibilità. Searle potrebbe sostenere che il manuale di istruzione contiene tutti i casi possibili di domande e risposte. Tale affermazione implicherebbe, però, l'abbandono del riconoscimento della produttività illimitata del linguaggio naturale.

[34] Lo stesso Searle nella 2) invita a considerare il modello di un cervello come un sistema di manipolazione simbolica di simboli binari. L'idea della binarietà, peraltro, non corrisponde alla realtà dei fatti neurofisiologici, ma con opportuni aggiustamenti anche il cervello può essere considerato un sistema di elaborazione formale di simboli. Bisogna, però, valutare approfonditamente fino a che punto tali aggiustamenti riescano a mantenerne le specifiche caratteristiche funzionali, questione ancora aperta all'interno delle neuroscienze cognitive.

[35] Reazioni a favore e contrarie, che nella rivista *Behavioral and Brain Sciences* sono riportate insieme al saggio di Searle.

anno dalla sua pubblicazione[36]. Essa costituisce una premessa utile alla discussione delle tesi hofstadteriane in merito alla metodologia e agli obiettivi dell'IA, su cui verterà il resto di questo scritto.

La risposta di Hofstadter (e Dennett) alla stanza cinese è quella "dei sistemi". Questo non sorprende, accettando come valida la riduzione delle sei obiezioni proposte da Searle all'unica che non consista in una fallacia argomentativa (nel senso di non essere direttamente rivolta all'argomento) o in una ripetizione delle obiezioni precedenti, vale a dire la 1). Di essa sono già stati messi in luce i punti deboli e le implausibilità (mascherate con robusti punti di forza intuitivi). Hofstadter, la cui critica dell'argomento di Searle sottolinea queste debolezze, richiama in aggiunta il meccanismo della "pompa di intuizione" che Dennett aveva introdotto proprio in riferimento all'argomento della stanza cinese. Una "pompa di intuizione" è «un congegno che provoca una serie di intuizioni col produrre variazioni su un esperimento di pensiero basico» (Dennett, 1980, p. 94). Tali variazioni permettono di ricavare dalla stessa struttura argomentativa conclusioni diverse a seconda delle caratteristiche attribuite ad un qualche *Gedankenexperiment* preso in considerazione. In riferimento alla stanza cinese, Hofstadter individua cinque parametri, «cinque manopole», sulla base dei quali è possibile variare la situazione ideale descritta dall'esperimento (Hofstadter, 1981, p. 363):

- il *materiale* fisico su cui viene costruita la simulazione;
- il *livello imitativo* del sistema mente-cervello (subatomico, atomico, sinaptico, cellulare neurale, di gruppi di neuroni, simbolico, ecc.);
- la *grandezza fisica* della simulazione (dal microscopico al macroscopico);
- la grandezza e la tipologia del *demone* della simulazione, cioè il principale attore della simulazione;
- la *velocità* d'azione del demone (molto lenta o molto veloce).

Queste cinque variabili rendono possibile la creazione di molteplici e differenti esperimenti della stanza in cui, generalmente, è presumibile che l'effetto intuitivo sia ottenuto mediante i semplici accorgimenti di rallentare notevolmente la velocità di esecuzione del compito, di ingrandire a dimensioni umanamente inconcepibili il sistema globale, di utilizzare materiale quanto più possibile inerte e inattivo, di introdurre all'interno della stanza, cioè del nucleo centrale dell'esperimento, un demone che sia il più simile possibile a un agente umano (e quindi anche un agente umano stesso), che compia meccanicamente, o comunque *metodicamente*, determinate azioni (pur potendo comportarsi in maniera non meccanica, essendo *human-like*) e che, effettivamente, *sostituisca la parte essenzialmente esplicativa* dell'esperimento di simulazione.

[36] Cfr. Hofstadter, Dennett (1981, pp. 360-369). Il commento a "Menti, cervelli e programmi", pur esprimendo convincimenti condivisi da entrambi gli autori, porta la firma di Hofstadter.

In questo modo, è possibile ottenere la stanza simbolica di Searle (la stanza dei simboli cinesi), ma anche la stanza subsimbolica di Haugeland[37] (la stanza delle connessioni sinaptiche), in cui il demone presente è capace di attivare, pizzicandole, tutte e sole le giuste sinapsi di un cervello all'interno di un individuo che conversa in cinese (Haugeland, 1980, pp. 108-109). Sappiamo già quale sia l'obiezione di Searle a questa trasformazione della stanza. Nella obiezione 3) egli afferma che anche in questo caso non viene meno la tesi principale, ovvero la mancanza di intenzionalità, e quindi di comprensione del cinese, da parte dell'individuo che muove le leve (idrauliche o elettriche) nella stanza cerebrale. Nel caso di Haugeland, però, Searle sembra ritrattare quanto detto nella 3), quando nelle risposte alle obiezioni egli afferma che una riproduzione così esatta del cervello, con la sostituzione di un demone al normale svolgimento dell'attività sinaptica, non cancella la presenza dei poteri causali del cervello, perché «se la stimolazione delle cause è a un livello abbastanza basso da riprodurre le cause e non semplicemente descriverle, la "simulazione riprodurrà gli effetti» (Searle, 1980, p. 198). Ora, delle due l'una: o fra la 3) e l'ammissione che un demone che si sostituisca in maniera perfetta alle interazioni sinaptiche di un cervello non c'è alcuna differenza, circostanza che equivale a una ritrattazione da parte di Searle della sua posizione iniziale; oppure si deve ammettere che l'unica differenza fra le due situazioni sia relativa al *materiale* con cui viene costruito il sistema in grado di attuare la comprensione del cinese: i poteri causali del cervello sono da considerare in ogni caso differenti dai poteri causali presenti all'interno di un complesso di tubi e valvole idrauliche (o di "pizzicatori" di sinapsi) che replica fedelmente la struttura di un cervello. Tuttavia, la natura di un potere causale, per definizione, non è individuata da una particolare proprietà di una determinata sostanza materiale, quanto piuttosto consiste nella possibilità stessa di essere specificato come la potenzialità di produrre certi effetti a partire da certe condizioni, anche, ma non solo e non necessariamente, materiali (si pensi ad esempio a una sostanza chimica che si trasforma in un'altra), in base a un determinato procedimento o metodo.

1.7 Il problema di Searle e il "ciclo di purificazione" dei modelli

Tuttavia, su tale questione Searle non arriva a dare un adeguato chiarimento, verosimilmente perché il *focus* della sua attenzione rimane esclusivamente quello del linguaggio naturale. Sia Dennett che Hofstadter pongono il problema di che cosa sia veramente apprendere una lingua diversa rispetto a quella che si parla come madrelingua, arrivando a concludere che non è possibile che l'individuo nella stanza possa "internalizzare" tutte le istruzioni del manuale, in modo da rendere non valida la 1), l'obiezione del sistema (individuo + istruzioni). Infatti, se "internalizzare" vuol dire memorizzare, è ancora sempre il sistema che comprende. Non c'è differenza tra l'avere qualcosa scritto su un foglio di carta o nel ricordarlo pedissequamente per come è scritto su quel

[37] Da lui introdotta nella risposta a Searle nello stesso numero di *Behavioral and Brain Sciences.*

foglio, come ha suggerito Putnam nel suo *Gedankenexperiment*. L'unica differenza sta nell'impiego di una gran quantità di risorse di memoria da parte del memorizzante. Se, al contrario, "internalizzare" il programma vuol dire inserirlo nei propri "sottosistemi" non si vede come questo possa essere fatto senza attuare una qualche forma di collegamento fra il programma e i sottosistemi, il quale dia luogo ad un uso consapevole delle conoscenze "internalizzate". Si tratterebbe, in conclusione, di apprendimento, e, in questo particolare caso, dell'apprendimento di un'altra lingua[38].

Ancora una volta, però, non è questo che probabilmente interessa Searle, o ciò che lui veramente intende con l'argomento della stanza cinese. Il problema centrale resta quello della simulazione della comprensione (e produzione) del linguaggio naturale e delle obiezioni che possono essere sollevate nei confronti di questa particolare attività cognitiva. I parametri di variazione dell'argomento individuati da Hofstadter possono essere visti, *mutatis mutandis*, come un'incompleta, ma efficace, lista di restrizioni a tutti i modelli simulativi, vale a dire, applicabili in linea di principio al retroscena teorico dei tentativi di simulazione di tutte le attività cognitive. Essi valgono anche nel momento in cui si affronta la comprensione, in un senso più estensivo di elaborazione del linguaggio naturale, come problema dell'IA. Con una differenza. Il linguaggio *di per sé* pone il problema di come debba essere considerato, fra i due estremi del puro episodio comportamentale esteriore, mero output di una serie di meccanismi, procedimenti, funzioni (simboliche o biologiche o entrambe) che si svolgono in un'interiorità costituita dalla mente e/o dal cervello, e della manipolazione simbolica in base a regole sintattico-formali e a regolarità semantiche che insieme permettono la comprensione e la produzione del linguaggio. L'analisi compiuta dell'argomento della stanza cinese ha mostrato come Searle si muova fra un estremo e l'altro, confondendoli e spingendo oltre limiti accettabili di plausibilità la situazione ideata da Turing nel gioco dell'imitazione.

Questo induce un'ultima riflessione. Si era parlato di un doppio passaggio che permettesse la costruzione del *Gedankenexperiment* della stanza cinese a partire dal gioco di Turing. Ora possiamo identificare meglio questa duplice trasformazione in due mosse specifiche. La prima, dal Gioco al Test, è una "*mossa comportamentistica*", che trasforma la natura del linguaggio naturale nella simulazione da mezzo di comunicazione a output di una determinata attività cognitiva a garanzia della effettiva presenza di quest'ultima all'interno della macchina simulativa. La seconda, dal Test alla Stanza, la quale deriva direttamente dalla natura simbolica della componente segnica, fonetica e grafica, del linguaggio naturale, è una "*mossa formalistica*", attraverso cui esso non è più soltanto un output di un'attività nella mente o nel cervello, ma qualcosa di interno a essi che può essere

[38] A questo punto, disquisire se apprendere una lingua attraverso un fantomatico manuale di istruzioni per rispondere a domande su un episodio narrato in quella lingua sia la stessa cosa che apprenderla attraverso un manuale di grammatica è lo stesso che chiedersi se c'è una differenza, non esclusivamente metodologica, tra chi apprende una lingua attraverso un corso teorico di insegnamento e chi, invece, a stretto contatto con la realtà sociale in cui quella lingua viene parlata. La diversità del risultato non sembra implicare l'implausibilità di nessuno dei due metodi.

ridotto a una serie di istruzioni le quali, allo stesso tempo, lo formalizzano e lo rendono impermeabile alla comprensione.

Tali mosse si possono applicare, separatamente, ai programmi dell'IA che hanno in qualche modo cercato di simulare differenti capacità cognitive. In altri termini, il rischio di una deriva comportamentistica interessa tutti i modelli dell'attività mentale, o di una qualche specifica attività mentale, anche quelli puramente connessionisti, nel momento in cui si verifica *l'identificazione della spiegazione di una prestazione con l'esecuzione della medesima*. D'altro canto, la riduzione di una prestazione a un procedimento che goda delle stesse caratteristiche di inesorabile formalità, meccanicità e rigidezza di una logica deduttiva (anche se predicativa e non "soltanto" proposizionale) è pure un punto di vista attraverso cui interpretare i differenti modelli dell'attività mentale, ma occorre che sia ben calibrato, per non ricadere in una prospettiva così analitica da perdere il suo potere esplicativo. Tale atteggiamento sembra risultare valido soltanto nella misura in cui viene considerato come uno dei punti di vista, necessario ma non sufficiente ai fini esplicativi, secondo cui valutare un modello dell'attività mentale. Nei casi del linguaggio naturale e della sua comprensione si è visto quanto facilmente siano soggetti a distorsioni dovute all'applicazione di queste due operazioni. Esiste una chiusura del circolo, una terza mossa che conduca nuovamente alla situazione iniziale, in un ciclo di verifica e filtrazione dalle obiezioni teorico-epistemologiche, il processo di costruzione dei modelli simulativi dell'IA? È un'ipotesi metodologica raffigurabile come in figura 1.1.

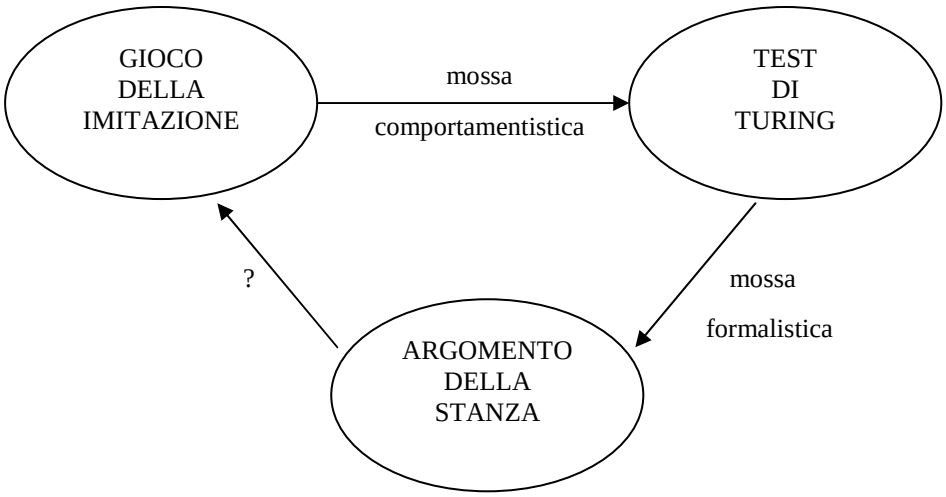

Fig. 1.1

Lo schema che ne deriva può essere considerato una sorta di "ciclo di purificazione" dei modelli, nel senso che, nel proporre simulazioni cognitive, e quindi nell'ipotizzare una qualche spiegazione di un processo o di un fenomeno mentale, è sempre opportuno considerare i livelli di comportamentismo e di formalismo presenti nella componente esplicativa della simulazione e trarne

le opportune conseguenze, anche in termini di revisione del modello o della teoria che lo supporta, qualora non venga prodotta sufficiente o effettiva spiegazione dell'attività cognitiva indagata.

Ma per quali ragioni, in senso specifico, dovrebbe essere auspicabile un ritorno a Turing e allo spirito del suo gioco dell'imitazione? Per due motivi almeno, legati entrambi al senso profondo del gioco da lui proposto, quello di valutare a che condizioni *noi sperimentatori* saremmo disposti ad ammettere di trovarci in presenza di macchine pensanti. In primo luogo, dal punto di vista metodologico. Come in altre discipline scientifiche, così anche nell'IA e nelle scienze cognitive non è mai conveniente sovrastimare la portata di un esperimento (simulativo). Occorre, invece, valutare attentamente il fenomeno in via di sperimentazione, fissarne le restrizioni, cioè *le condizioni a cui quel fenomeno continua a rimanere quel dato fenomeno anche nella simulazione*, e infine anticipare e verificare i risultati attesi. In secondo luogo, dal punto di vista teorico ed epistemologico. Infatti, bisogna avere una chiara idea del fenomeno che si intende modellare e non trascurare mai il fatto che il legame con la realtà del modello, almeno e necessariamente per qualche aspetto, non deve essere frutto di un'attribuzione dall'esterno, cioè da parte di un osservatore, fatto che esporrebbe inevitabilmente il modello alle critiche evidenziate in precedenza. Ne consegue che l'ultima mossa, quella del ritorno, si configura come una "*mossa realistica*" e il suo intento costituisce un richiamo a un imperativo epistemologico che lo studio dei processi di pensiero attraverso metodologie simulative non può disattendere, pure nella provocatoria circostanza, di cui si diceva all'inizio, che tali metodologie costituiscano una via intermedia di sperimentazione dei fenomeni oggetto della loro indagine.

1.8 Leibniz e il mulino della percezione

Il "ciclo di purificazione" introdotto nel paragrafo precedente descrive un possibile schema di valutazione epistemologica del processo di progettazione dei modelli computazionali, ovvero di quel processo che va dalla teoria alla realizzazione del modello. Esso evidenzia, fra le altre cose, l'importanza del ruolo ricoperto dal linguaggio *in quanto sistema di simboli e di relazioni fra essi* dal punto di vista dell'epistemologia dell'IA e delle scienze cognitive. Come si è affermato in precedenza, infatti, la questione della comprensione (e produzione) del linguaggio naturale costituisce, ad esempio, un tema cardine dell'IA e mette in evidenza meglio di altri, nella sua ambiguità e complessità, i problemi relativi alla costruzione di modelli simulativi ed esplicativi di fenomeni mentali (o cerebro-mentali). Per concludere questa esposizione sull'argomento della stanza, vedremo come già in età moderna esso fosse stato applicato a un'altra attività mentale di alto livello: la percezione[39].

[39] Al contrario della sensazione che può essere considerata un'attività mentale di basso livello. Naturalmente, si tratta di etichette descrittive avalutative, che servono solo a distinguere una presunta, ma tradizionalmente ben consolidata e

Nel 1714 Leibniz scrive la *Monadologia*, che costituisce una summa sistematica del suo pensiero insieme ai *Principi razionali della Natura e della Grazia*. In quell'opera egli presenta il seguente argomento:

> Si deve riconoscere che la *percezione*, e quel che ne dipende, è *inesplicabile mediante ragioni meccaniche*, cioè mediante le figure e i movimenti. Immaginiamo una macchina strutturata in modo tale che sia capace di pensare, di sentire, di avere percezioni; supponiamola ora ingrandita, con le stesse proporzioni, in modo che vi si possa entrare come in un mulino. Fatto ciò, visitando la macchina al suo interno, troveremo sempre e soltanto pezzi che si spingono a vicenda, ma nulla che sia in grado di spiegare una percezione. Quindi la [ragione della] percezione va cercata nella sostanza semplice, non già nel Composto, cioè nella macchina. Così è unicamente nella sostanza semplice che si possono trovare le percezioni e i loro mutamenti: solo in ciò, quindi, possono consistere tutte le *azioni interne* delle sostanze semplici. (Leibniz, 1714/2001, p.65)

Questo argomento è stato variamente interpretato nella letteratura come un argomento antiriduzionista e antinaturalista[40]. Di conseguenza, si è sostenuto che con esso Leibniz abbia voluto negare la possibilità dell'attribuzione di stati mentali a stati fisici. Per Churchland, ad esempio, Leibniz non sa dove guardare, perché le sue conoscenze in merito al cervello sono inadeguate per indicare quali meccanismi neuronali possano realizzare la percezione, e, inoltre, in che modo possano farlo. Ciò lo porterebbe a una negazione del fenomeno, piuttosto che al riconoscimento di un'ignoranza contingente in merito e relativa allo stato delle conoscenze scientifiche raggiunte. Un neurofisiologo contemporaneo ha – o comunque avrà entro un certo periodo determinato di tempo – sicuramente gioco facile nell'individuare il fenomeno fisico cui può essere ridotta la percezione (Churchland, 1995, pp. 191-193).

Per Searle si tratta, più semplicemente, di una confusione dei livelli di descrizione (Searle, 1983, pp. 268-273). Eventi mentali causano eventi mentali, così come eventi fisici causano eventi fisici. Ma anche, eventi fisici realizzano, e perciò causano, eventi mentali. Di conseguenza, se un evento fisico realizza (causa) un evento mentale che causa (realizza?) un altro evento mentale, per la proprietà transitiva della causazione l'evento fisico primo è causa (anche) dell'ultimo evento mentale. Cercare, però, il mentale *nel* fisico senza un'adeguata conoscenza di come l'uno si riduca all'altro è una confusione di livelli che ha come diretta conseguenza il paradosso della negazione del fenomeno o che, più verosimilmente nell'ottica di Leibniz, porta a un riconoscimento dell'esistenza di una differenza ontologica fra i due livelli. Ma, anche per Searle, tutto ciò è solo questione di ignoranza: «se avessimo una conoscenza perfetta di come il cervello produca sete o esperienze visive, non avremmo nessuna esitazione nell'assegnare queste collocazioni di esperienza

---

accettata nelle scienze cognitive, distanza maggiore o minore dal cervello e dal livello neurofisiologico di indagine, ovvero anche dall'ambiente in cui è immerso il sistema cognitivo che agisce e percepisce.

[40] Per una rassegna delle critiche all'argomento si veda Calabi (2005, p. 194).

nel cervello, se l'evidenza garantisse questi assegnamenti» (Searle, 1983, p. 271), e questo varrebbe anche nel senso di una localizzazione globale di eventi mentali in tutto il cervello o in vaste aree di esso. Ancora una volta, Leibniz mancherebbe di riconoscere la riconducibilità ultima del mentale ai poteri causali del cervello.

Ma come avrebbe potuto? Nel 1714, anno in cui viene redatta la *Monadologia*, il paradigma dualista inaugurato da Cartesio con il riconoscimento di due sostanze separate a comporre per giustapposizione l'unità dell'essere umano è all'apice della diffusione e del consolidamento. Anche per Leibniz il dualismo fra anima (mente) e corpo, o fra pensiero in quanto cosa pensante (*res cogitans*) e substrato materiale in quanto cosa estesa (*res extensa*), è un dato di fatto e allo stesso tempo un problema risolvibile soltanto attraverso l'armonia prestabilita la cui comprensione trascende l'ambito del mondo fisico. Di conseguenza, non può che essere connaturato con la totalità del suo sistema, in maniera radicale e indubitabile, l'assunto di una differenza di stampo ontologico fra stati fisici e stati mentali. Non deve stupire, perciò, che l'argomento del mulino conduca a esiti antiriduzionisti. Tuttavia, non ritengo che Leibniz lo abbia formulato con questo intento. Più plausibile sembra, invece, l'assegnare a esso un ruolo centrale nella definizione di un tipo accettabile di spiegazione dei fenomeni mentali e dei processi cognitivi.

Riconsideriamo l'argomento. L'attenzione di Leibniz appare essere tutta rivolta a quelle *ragioni meccaniche* mediante cui è *inesplicabile* un fenomeno come la percezione. Questo porta a pensare non che la percezione sia un fenomeno inspiegabile, bensì che ci sia un qualche altro tipo di spiegazione possibile, che si diano cioè «due tipi di spiegazione» per i fenomeni mentali, le qual,i come suggerisce Calabi, sono «la spiegazione per ragioni meccaniche e la spiegazione naturale» (Calabi, 2005, p. 194). Delle due, la prima sarebbe propriamente una spiegazione riduzionistica, e perciò finita e incompleta; la seconda «è una spiegazione che fa riferimento alle cause finali e non alle ragioni sufficienti e, in ultima analisi, equivale a una spiegazione per ragioni meccaniche che è infinitamente lunga» (*ibidem*). A partire da questa interpretazione Calabi ipotizza che Leibniz non introduca l'argomento del mulino per arrivare a conclusioni ontologiche in merito ai fenomeni mentali e conclude che Leibniz non era un riduzionista concettuale, ma piuttosto un riduzionista metafisico. L'insufficienza esplicativa sarebbe dovuta al fatto che la spiegazione naturale richiede un'analisi infinita, fattualmente impossibile, e l'intera questione si risolve, anche per Leibniz stesso, in un'indecidibilità in merito alla questione se gli stati mentali sono o non sono (riducibili a) stati fisici.

D'altra parte, se si accetta l'idea che il *Gedankenexperiment* del mulino «non è un argomento che da premesse epistemologiche conduce a conseguenze ontologiche» (Calabi, 2005, p. 210) ed è verosimile, come ho sostenuto, che l'intento di Leibniz non era quello di introdurre un argomento antitiduzionista in merito alla natura degli stati mentali, non è del tutto forzoso vedere nella situazione descritta da Leibniz non un rimando a una spiegazione soltanto di tipo metafisico della percezione, bensì l'affermazione che il meccanicismo inteso nel senso di una serie di interazioni

sequenziali causa-effetto non può essere considerato una spiegazione completa senza la sua integrazione con una visione di tipo finalistico, o relativa alle cause finali, del fenomeno stesso della percezione. In altri termini, Leibniz starebbe suggerendo la "risposta del sistema". Vediamo in che modo è possibile argomentare questo punto.

L'esperimento mentale del mulino prende l'avvio dall'ipotesi di una «macchina strutturata in modo tale che sia capace di pensare, di sentire, di avere percezioni». Entrare in tale macchina (la stanza con i macchinari del mulino) ci permette di vedere «sempre e soltanto pezzi che si spingono a vicenda». Tuttavia, per Leibniz tale macchina esiste, cioè esiste una macchina in grado di pensare e percepire *grazie alla sua struttura*. Di che macchina si tratta?

Nel paragrafo 64 della *Monadologia*, dopo aver già introdotto l'argomento del mulino, egli afferma che «il corpo organico di ogni essere vivente è una specie di macchina divina, o di automa naturale, che supera di gran lunga qualsiasi automa artificiale», nel senso che, rispetto alle macchine costruite dall'uomo «le macchine della Natura, cioè i corpi viventi, sono sempre delle macchine, fin nelle loro parti più minute, all'infinito» (Leibniz, 1714/2001, p.89). Il corpo umano è, dunque, una macchina i cui pezzi sono ancora delle macchine, mentre le macchine costruite dall'uomo sono costituite da «parti o frammenti che per noi non sono più qualcosa di artificiale e che, riguardo all'uso cui [la macchina] è destinata, non serbano più nessuna traccia meccanica» (*ibidem*). Questo suggerisce un'idea del corpo vivente come di una serie gerarchica di macchine, analizzabili ciascuna in quella di livello immediatamente inferiore, senza la possibilità di arrivare mai a un livello base. Si potrebbe vedere adombrata in queste affermazioni la moderna differenza fra genotipo e fenotipo, con l'importante differenza che nell'ipotesi di Leibniz non esiste una base genetica ultima. Tuttavia, il passaggio da un livello a quello superiore è dovuto, di volta in volta, alla presenza di una differente struttura organizzativa che caratterizza il livello in oggetto. Leibniz ci dice, inoltre, che anche l'anima è un automa meccanico e precisamente «un automa immateriale, la cui costituzione interna è una concentrazione o rappresentazione di un automa materiale, e produce, rappresentativamente, in questa anima lo stesso effetto» (Leibniz, 1963, p. 280). È questa la macchina che ci interessa, poiché, se la differenza ontologica, il dualismo delle sostanze, caratterizza la differenza fra automa naturale e automa immateriale, ciò che tra le due sostanze si mantiene è proprio lo stesso concetto di meccanicismo, applicabile, nello stesso tempo e alla stessa maniera, ad entrambi gli automi. Infatti, sia gli automi naturali che quelli immateriali contengono una loro *peculiare struttura* per via della preformazione divina che li ha creati e li ha messi in condizione di operare meccanicamente, seppur su piani differenti: «l'operazione degli automi spirituali, vale a dire delle anime, non è meccanica, bensì contiene eminentemente quanto vi è di bello nella meccanica» (Leibniz, 1710/2000, p. 388). *L'automa spirituale è, perciò, la rappresentazione dell'automa materiale, la rappresentazione della sua meccanicità secondo un principio di unità, che è quello della monade*.

Per Leibniz la rappresentazione ha un ruolo centrale, non diversamente dai filosofi che nel diciassettesimo secolo e ancora negli anni in cui egli scriveva si occupavano di filosofia della conoscenza. La rappresentazione non è altro che la percezione stessa[41], la quale non può darsi, cioè spiegarsi, nella scomposizione delle sue parti, o, meglio, nelle parti della macchina che la producono, ma risiede nel principio della sua unità, che è la monade, sostanza semplice e automa immateriale su cui si riflette la meccanicità delle parti del corpo materiale. Così si ritorna alla conclusione dell'argomento del mulino. Il principio di unità, alleggerito dal suo bagaglio ontologico, cioè a prescindere dalla inconoscibilità della sua metafisica natura ultima, può essere non avventatamente considerato principio di organizzazione strutturale. Di conseguenza, se si dà una macchina in grado di pensare e percepire, come Leibniz afferma, e vi si entra, non si vedrà nulla all'infuori di parti meccaniche che ne spingono altre[42], a meno che non si conosca la funzione di ogni parte, le relazioni che legano le varie parti e l'organizzazione globale di tutto il sistema. Se non si accetta il principio esplicativo della struttura organizzativa – che si usi o meno una terminologia finalistica –, non si vede a che cosa possa servire nella situazione descritta da Leibniz la presenza di una macchina. Se la percezione risiedesse solo nella monade come principio trascendente, l'argomento del mulino sarebbe la negazione assoluta del meccanicismo, il che contrasterebbe con l'affermazione, pur non del tutto chiara, di Leibniz relativa all'attività degli automi spirituali, la quale non è meccanica, ma contiene ciò che di eminentemente bello è presente nella meccanica, ovvero la struttura globale relazionale e l'organizzazione funzionale unitaria delle parti connesse secondo leggi di causa-effetto[43].

1.9 La stanza fra linguaggio e percezione

L'interpretazione che si è data dell'argomento di Leibniz vuole essere coerente con l'idea che non bastano le leggi causali che governano le componenti del cervello a spiegare determinati fenomeno cognitivi. Come nel caso degli esperimenti mentali presi in considerazione più sopra,

[41] Che le percezioni, cioè le rappresentazioni, abbiano un ruolo funzionale nello svolgimento del pensiero, nel senso proprio di veicolare informazioni e di causare altre rappresentazioni, è questione che oltrepassa i limiti di questo discorso. Per rimanervi dentro, basti considerare che l'argomento del mulino riguarda essenzialmente il modo in cui un individuo può affermare di *avere o possedere una particolare percezione*.

[42] Indipendentemente dal fatto che essere costituiscano o meno a loro volta macchine ulteriormente analizzabili. Nella visione di Leibniz non c'è nulla che vada contro l'ipotesi secondo la quale ogni livello *contiene*, in relazione a quello immediatamente inferiore, le proprie ragioni esplicative meccaniche e strutturali.

[43] Un altro indizio a favore di questa intepretazione può essere visto in quelle *petites perceptions* di cui Leibniz parla nella Prefazione ai *Nuovi saggi sull'intelletto umano* e che determinano, *in modo inconscio* e, a quanto si può capire, attraverso una modalità interattiva complessa, la percezione cosciente: «Queste piccole percezioni sono [...] di più grande efficacia di quanto si pensi. Sono esse che formano questo non so che, questi gusti, queste immagini delle qualità dei sensi, chiare nell'insieme, ma confuse nelle parti; queste impressioni che i corpi circostanti producono e che racchiudono l'infinito, questo legame che ciascun essere ha con tutto il resto dell'universo. Si può anche dire che, in conseguenza di queste piccole percezioni, il presente è pieno dell'avvenire e carico del passato, che tutto è conspirante [...]» (Leibniz, 1705/1982, p. 49).

anche qui ci troviamo di fronte a una situazione in cui variazioni apposite producono uno stato di cose solo idealmente esperibile, il quale può essere considerato, *mutatis mutandis*, come l'esperienza di visualizzare singoli neuroni, o insiemi di neuroni di grandezza crescente, che scaricano a una velocità rallentata fino al punto che possiamo percepirli attraverso i nostri sensi nell'atto di produrre la percezione. Con la sua operazione argomentativa, Leibniz vuole veicolare l'idea che conoscere il modo in cui i singoli pezzi si muovono, le leggi dell'eccitazione e della scarica, non è sufficiente per una spiegazione completa di questo fenomeno mentale. Occorre avere anche una visione d'insieme dell'intero meccanismo. Questa è la risposta del sistema, che sembra implausibile fino a che si consideri il sistema, come fa Searle, soltanto una mera giustapposizione di parti differenti[44] e non una struttura relazionale organizzata.

L'avere affrontato il tema della percezione in relazione all'argomento della stanza ci permette un'ultima riflessione. È abbastanza evidente che le due mosse cui si è accennato in precedenza non sono applicabili alla situazione immaginata da Leibniz. Non si può avere una prova esteriore dell'avere una percezione, se non attraverso un resoconto linguistico o una constatazione operativa delle azioni e del comportamento del sistema. Non è tanto questo, però, che Leibniz intende sottolineare con il suo argomento, quanto piuttosto individuare il livello adeguato e le categorie concettuali adatte per poter dare un resoconto esplicativo del fenomeno. L'aspetto formalistico non viene preso in considerazione, né potrebbe esserlo, a meno che non si voglia attribuire un significato formale alle "ragioni meccaniche" che governano il movimento dei pezzi del mulino. La stanza di Leibniz differisce da quella di Searle, e, dunque, differiscono anche le conclusioni che se possono trarre in merito ai due fenomeni coinvolti. Con quali conseguenze?

Nella sua analisi dell'argomento della stanza cinese Chalmers arriva a esiti analoghi a quelli cui siamo giunti in queste pagine, vale a dire che non siamo in presenza di un argomento stringente contro il computazionalismo *tout court* e la possibilità di implementare la comprensione e i fenomeni coscienti in generale (Chalmers, 1996, p. 332). Tuttavia, è proprio il generalizzare l'argomento a qualunque tipo di esperienza cosciente[45] a non sembrare attuabile. L'argomento della stanza cinese funziona se è sotto esame la comprensione linguistica, e solo quella, in quanto capacità cognitiva da implementare in un programma. Questo è dovuto al fatto che solo il linguaggio, o, meglio, una specifica lingua culturalmente e storicamente determinata quale insieme di simboli fonetici e grafici, si presta all'operazione effettuata Searle per comprovare e rafforzare le sue tesi contro il computazionalismo dei fenomeni mentali. Naturalmente questo vale *a fortiori* per qualsiasi linguaggio formale specifico, ma l'argomento rappresenterebbe in questo caso una

[44] «L'idea è che, mentre una persona non comprende il cinese, in qualche modo la *combinazione* di quella persona e di pezzi di carta potrebbero, insieme, capire il cinese: non è facile per me immaginare che qualcuno (che non fosse nella stretta di un'ideologia) potrebbe trovare l'idea in qualche modo plausibile» (Searle, 1980, p. 53). In qualunque modo si voglia valutare questa affermazione, rimane il fatto che il suo grado di plausibilità è esattamente lo stesso dell'argomento della stanza cinese.

[45] «Si prenda un programma che è supposto catturare *qualche aspetto della coscienza*, come comprendere il cinese o avere la sensazione di rosso» (Chalmers, 1996, p. 329, [corsivo mio]).

situazione banale. È ovvio, infatti, che un linguaggio formale possa essere trattato in maniera meccanica; é così per definizione. Non è ugualmente chiaro in che modo e fino a che punto è possibile trattare meccanicamente il linguaggio naturale e proprio da questa discrepanza trae forza l'argomento della stanza cinese. In altri termini, ciò che è sotto indagine è il grado di formalismo, cui un linguaggio naturale deve essere ridotto o con cui deve essere analizzato, per poter essere implementato *meccanicamente*. Questo appare anche più evidente se si considera che gli stessi linguaggi di programmazione sono linguaggi formali ai quali si adatterebbe molto di più un trattamento simile a quello riservato al cinese all'interno della stanza, anche se in misura sempre minore man mano che si risale la scala gerarchica dei linguaggi da quello macchina fino a quello naturale, utilizzato per la formulazione della pre-struttura algoritmica di un programma.

Chalmers stesso, nell'esposizione della sua versione dell'argomento, sembra confermare il fatto che la stanza cinese funzioni soltanto se riferita alla comprensione del linguaggio e non a qualunque aspetto della coscienza. Nella sua descrizione, infatti, ripropone lo schema originario di Searle, adoperando come esempio "paradigmatico" la (non) comprensione della lingua cinese. In realtà, non c'è nulla di paradigmatico, bensì si tratta di un'esclusività dovuta all'effettivo darsi, di volta in volta in una forma concreta parlata o scritta, della *natura squisitamente simbolica* del linguaggio, una forma che è storicamente, socialmente o convenzionalmente – si pensi ai linguaggi formali e ai linguaggi di programmazione – determinata. Perciò, al di là della disputa se questo argomento si riferisca soltanto all'intenzionalità o anche alla coscienza, che qui ci interessa solo marginalmente, mi pare che il modo in cui Charlmers ricostruisca l'argomentazione mostra che non si possono mettere sullo stesso piano linguaggio e percezione (intesa come categorizzazione di una sensazione) e che c'è una profonda differenza fra la stanza cinese e il mulino senziente descritto da Leibniz: il primo è un argomento contro una spiegazione formalistica, che però non esaurisce tutte le forme di computazionalismo, dell'attività cognitiva, mentre il secondo è un *caveat* nei confronti del corretto atteggiamento esplicativo da impiegare nel dare resoconti dei processi del pensiero.

La tendenza ad assimilare l'IA e le scienze cognitive precedenti l'affermazione del punto di vista connessionista ad una ricerca che ha come esclusivo costituente del pensiero una concezione modulare e sintattico-manipolativa dei contenuti della mente è a metà strada tra l'essere adeguatamente e approssimativamente realistica. Di certo, l'IA dei primi quaranta anni non può essere ridotta soltanto allo studio del Linguaggio del Pensiero, delineato da Fodor. Tuttavia, questo ne è stata una componente fondamentale. La versione forte del computazionalismo sottesa alla modularità della mente e al Linguaggio del Pensiero ha influenzato profondamente il campo di studi delle scienze cognitive, in alcuni casi apportando benefici proprio attraverso l'affermazione dogmatica e perciò provocatoria di questi due assunti teorici. In particolare, l'idea di un Linguaggio del Pensiero ha posto l'attenzione sul ruolo centrale che hanno i *concetti*, interpretati come "le parole del pensiero" soprattutto a causa della plausibilità di una loro connotazione unitaria e stabile *proprio come una parola* del linguaggio naturale, in attività cognitive, quali la memoria, la

produzione del linguaggio, l'apprendimento, la percezione. Le critiche a questa concezione hanno favorito la nascita di una nuova impostazione di ricerca nelle scienze cognitive, che dopo pochi decenni lascia ancora aperti numerosi problemi relativi alle attività cognitive di alto livello, sia per quanto riguarda la nozione di rappresentazione, sia per quanto riguarda la spiegazione di fenomeni come la percezione e la produzione e comprensione del linguaggio.

Nei prossimi capitoli la nostra attenzione si volgerà, perciò, a un'impostazione della ricerca all'interno delle scienze cognitive che non si propone di eliminare del tutto la parte simbolica del pensiero, ma che attua una profonda revisione nel proporre un differente approccio alla modellistica computazionale cognitiva, indagando aspetti tradizionalmente lasciati da parte dall'IA simbolica, quelli subcognitivi, e mettendo al centro la questione della rappresentazione della conoscenza e della modellizzazione dei *concetti*. Le riflessioni proposte in questo capitolo saranno il puntello d'appoggio in questo percorso per arrivare a nuove riflessioni epistemologiche nella parte finale di questo lavoro, tenendo ben presente che, ogni volta che il tentativo è quello di produrre un sapere scientifico e oggettivo, la validità di certe obiezioni non può essere cancellata attraverso il semplice stravolgimento degli obiettivi e il cambiamento delle metodologie impiegate. Questo vale anche, e in special modo, per tutte quelle discipline scientifiche che, volentieri o meno, devono fare i conti con la scomoda e ingombrante nozione di "mente".

Capitolo 2

L'APPROCCIO SUBCOGNITIVO ALL'INTELLIGENZA ARTIFICIALE

2.1 I principi della subcognizione

Per superare l'*impasse* scaturita dalle molteplici obiezioni rivolte all'IA simbolica nel corso degli anni '70, fra le quali quella di Searle svolge un ruolo cruciale, nuovi approcci sono stati proposti a partire dall'inizio degli anni '80. Il più influente nei decenni a venire è stato sicuramente quello connessionista, che ha spostato ad un livello diverso rispetto a quello simbolico l'implementazione dell'elaborazione, con, tra l'altro, riflessi cospicui sul modo di intendere filosoficamente il rapporto fra mente e cervello e quello fra meccanismi di pensiero e meccanismi di elaborazione.

L'approccio connessionista, anche a voler semplificare, non può essere considerato unitario e molteplici acquisizioni in questo campo si sono susseguite negli anni, sia dal punto di vista della crescente complessità delle reti neurali, che costituiscono l'aspetto implementativo per eccellenza dell'approccio connessionista, sia dal punto di vista degli scopi prefissati e conseguiti da questo filone di ricerca[1]. Il connessionismo, d'altra parte, non esaurisce la totalità degli approcci all'IA proposti negli ultimi vent'anni, anche se coglie, anzi si fonda su, uno dei tratti principali del nuovo modo di condurre la ricerca nel campo delle scienze cognitive: lo spostamento a un livello non simbolico esplicito dell'elaborazione dell'informazione.

Questa impostazione è condivisa in parte anche dall'approccio subcognitivo alla cognizione[2], il quale, però, ipotizza che il livello a cui deve essere condotta l'analisi e la spiegazione dei meccanismi del pensiero sia non quello neurale, come fa buona parte del connessionismo, ma quello *concettuale pre-simbolico*. In altri termini, si assume che il pensiero non vada trattato come mera

[1] Per un'introduzione particolareggiata ai presupposti teorici, alle metodologie e alle tecniche dell'approccio connessionista si rimanda a Floreano, Mattiussi (2002).

[2] Il termine "subcognizione" viene a volte utilizzato indifferentemente al posto di connessionismo. In questa sede ci sembra opportuno distinguere "subcognizione" da "connessionismo", in considerazione del fatto che questi due termini esprimono un diverso approccio al problema della rappresentazione in particolare e del sistema mente-cervello in generale. Infatti, mentre l'utilizzo di reti neurali in generale è strettamente collegato ad un prospettiva subsimbolica, o che si potrebbe anche definire a-simbolica, ed eliminativista, con tutte le ricadute problematiche nei confronti della simulazione e della spiegazione dei processi mentali di alto livello, l'approccio subcognitivo è ancora un approccio simbolico che sfrutta *soltanto in senso funzionale e architetturale, e non rappresentazionale*, alcune caratteristiche del cervello fatte proprie, sia metafisicamente che epistemologicamente, dalla metodologia connessionista.

elaborazione formale e sintattica di simboli, come suggerisce la teoria computazional-rappresentazionale della mente proposta da Fodor[3], bensì come il prodotto di una aggregazione di concetti (rappresentati) su molteplici livelli, la cui esplicitazione linguistica è soltanto uno degli aspetti derivati, anche se forse uno dei più difficile da spiegare all'interno di questa impostazione di ricerca. Nell'approccio subcognitivo il linguaggio diviene, si può dire, una sorta di finestra aperta sull'attività mentale alla cui base stanno i *concetti* concepiti come *entità funzional-causali* in grado di produrre quella forma sofisticata e complessa di ragionamento associativo che è il fare analogie e che soltanto per alcuni aspetti è riconducibile all'associazionismo della tradizione filosofica empirista[4].

Nei successivi capitoli si esporranno i prodotti più significativi di IA che rientrano in qualche misura in questo orientamento. L'esposizione e la valutazione dei modelli cognitivi conformi a questa impostazione proposti negli ultimi venti anni dovrebbe chiarire la portata e i limiti dell'approccio subcognitivo al mentale e schiudere la strada alle sue future prospettive. Alcuni dei modelli qui discussi sono stati già delineati, in maniera più o meno approfondita, in Hofstadter & FARG (1995). Alcuni passi avanti nel corso degli ultimi anni sono stati fatti dal gruppo di ricerca che si dedica a implementare modelli di questo tipo, il FARG (*Fluid Analogies Research Group*). L'esposizione dei modelli, perciò, riprende in parte e arricchisce quella del 1995 con l'aggiunta del lavoro compiuto nell'ultimo decennio. Come filo conduttore dell'esposizione si è scelto di utilizzare i domini in cui essi operano, per ragioni che saranno spiegate in seguito. Per ora basti dire che, abbastanza intuitivamente, è proprio nel loro rapporto con il "mondo reale" che in genere i prodotti dell'IA e delle scienze cognitive hanno incontrato le maggiori difficoltà e i più grandi ostacoli, e in merito ad esso sono state formulate le critiche di maggiore impatto sull'evoluzione della ricerca stessa.

A ulteriore chiarimento del modo in cui la teoria subcognitiva del mentale è stata implementata verranno presentati in questo capitolo le caratteristiche principali di questo approccio all'IA unitamente alla presentazione dei programmi che lo hanno ispirato: i modelli HEARSAY e HEARSAY II. I modelli cognitivi sviluppati dal gruppo di ricerca sui concetti fluidi (FARG) condividono tre aspetti caratteristici, uno rivolto agli scopi, uno ai contesti e uno al tipo di architettura cognitiva funzionale utilizzata. Essi sono, rispettivamente:

1. la simulazione dei meccanismi del pensiero umano coinvolti nella produzione di analogie;
2. la focalizzazione su *microdomini*;

[3] Su questo si veda Fodor (1976). Va comunque ricordato che le opinioni di Fodor in merito alla teoria da lui formulata sono andate incontro a variazioni nei decenni successivi.

[4] Non menzioniamo neppure l'associazionismo psicologico, tipico del comportamentismo, proprio perché le associazioni nei modelli subcognitivi riguardano il piano concettuale e non coppie associative stimolo-risposta alla base, ad esempio, della teoria dell'apprendimento di Thorndike o associazioni fra stimoli come nella teoria della memorizzazione di Ebbinghaus (cfr. Legrenzi, 1999). Questo mancato riferimento può essere visto come un'ulteriore indicazione delle divergenze fra approccio subcognitivo e approccio connessionista.

3. l’utilizzo di una strategia di ricerca stocastica e parallela.

Il fine di questa trattazione sarà quello di rendere espliciti i termini, le potenzialità e gli eventuali limiti di quella che è l’idea guida alla base dell’approccio subcognitivo ai meccanismi della mente, approccio secondo il quale per la comprensione e la spiegazione di come funziona la mente, almeno per quanto riguarda gli aspetti semantici, è rilevante ciò che ricade immediatamente sotto la soglia della percezione cosciente. In particolare, di contro all’affermazione di Herbert Simon in merito all’inutilità di indagare i processi mentali che ricadono sotto la soglia dei cento millisecondi, individuati da Simon nei processi di riconoscimento categoriale di stimoli familiari (Simon, 1981), Hofstadter postula che sono proprio i processi, «microscopici e paralleli», immediatamente precedenti il riconoscimento cosciente ad essere importanti dal punto di vista esplicativo (Hofstadter, 1983a, p. 161). L’interazione di un numero elevato di tali processi produce la cognizione, intesa come ascrizione categoriale ottenuta anche attraverso processi di mescolanza concettuale.

Tale prospettiva consegue da una rivalutazione del fenomeno della percezione nel campo della scienze cognitive, che si avvia negli anni Settanta del secolo scorso, e conduce a un’affermazione della sua importanza nei primi anni Ottanta in sede sperimentale di simulazione dei processi del pensiero attraverso l’implementazione di programmi di IA. Così si esprime Hofstadter al riguardo:

> Per me, il punto cruciale dell’Intelligenza Artificiale è questo: “Che cosa mai rende possibile la trasformazione di 100.000.000 di punti della retina in una singola parola “madre” in un decimo di secondo?” La percezione è tutta qui. (Hofstadter, 1985c, p. 633)

Il tentativo di arrivare a una simulazione dei processi percettivi, che caratterizza in modi differenti la ricerca in IA a partire dagli anni Ottanta in maniera sostanziale e diversa rispetto alle ricerche degli anni precedenti, ha avuto esiti alterni. In effetti, molti modelli connessionisti sono riusciti a produrre buoni risultati in questo campo. Tuttavia, riecheggiando la distinzione kantiana nel processo conoscitivo fra un’estetica trascendentale e un’analitica trascendentale, cioè fra intuizione e concettualizzazione, si può suddividere la percezione di cui l’IA si occupa in due tipologie distinte: la percezione di basso livello, che corrisponde a compiti di elaborazione del mero dato sensoriale, che può avere come risultato finale l’individuazione di un oggetto attraverso la sua ascrizione categoriale, cioè la sua inclusione in una classe (la “madre” che ci è dato di cogliere attraverso i sensi)[5], e la percezione di alto livello, che corrisponde al compito di estrazione del significato, nel senso dell’operazione di concettualizzazione di situazioni che implicano un elevato grado di astrazione.

[5] Se nella citazione si fa l’esempio del concetto “madre”, bisogna dire che, di fatto, le ricerche che si sono indirizzate allo studio della percezione di basso livello hanno scelto categorie più concrete cui ricondurre il dato percettivo. L’esempio principale sono gli studi sulla percezione di visiva di Marr (1982).

Seppure fra le due non esista una separazione netta, ma si dispiegano entrambe lungo un unico spettro che va dal semplice al complesso o, se si vuole, dal concreto all'astratto, il secondo tipo di percezione appare più intrinsecamente connesso con la struttura fondamentale dei meccanismi del pensiero. E proprio la simulazione della percezione di alto livello costituisce l'obiettivo fondamentale dei programmi che ricadono all'interno dell'approccio subcognitivo. Essa esprime il tentativo di superamento teorico dell'impasse prodottasi all'interno dell'IA tradizionale e simbolica già durante gli anni Settanta e che viene imprescindibilmente colto da Searle con il *Gedankenexperiment* della stanza cinese. Il vero bersaglio delle sue affermazioni sono da considerarsi, non semplicemente i programmi che comprendono il linguaggio naturale, ma i programmi che si avvalgono in maniera troppo disinvolta di un apparato simbolico la cui interpretazione viene lasciata al programmatore o all'utente. L'uscita dal "fomalismo" e dal sintatticismo della stanza non deve, però, necessariamente configurarsi come un'uscita dalla stanza, cioè come rinuncia alla spiegazione dei meccanismi del pensiero in quanto tali. Essi vanni ripensati, e, per così dire, *riprogrammati* su un effettivo standard esplicativo, come *meccanismi interpretativi attivi*, in grado di produrre, invece che darla per scontato, l'unità dei due momenti in cui consiste il fenomeno percettivo-cognitivo. La cognizione non può essere scissa dalla percezione. Piuttosto i due processi vanno visti in stretta simbiosi e compito dei sistemi che si vogliono definire intelligenti è quello di cogliere e mettere in pratica questa reciproca compenetrazione. Il fare analogie costituisce il punto esatto della loro convergenza.

2.2 La percezione come analogia

L'identificazione del processo di percezione con un processo di creazione di analogie riguarda quella che viene definita la percezione di alto livello. Secondo Chalmers, French e Hofstadter tale tipo di percezione si ha «a un livello di elaborazione in cui *i concetti cominciano ad avere un ruolo importante*» (Chalmers, French, Hofstadter, 1992, p. 187 [enfasi mia]). Essa comprende uno spettro che va dal concreto all'astratto, dal riconoscimento degli oggetti in un campo percettivo (casa, cane, fiore), alla comprensione delle relazioni (fuori, a destra di), all'elaborazione di situazioni più complesse (un sistema politico, la vita di un uomo, lo stile di un artista). Visto il ruolo ricoperto dalla conoscenza già codificata in questo tipo di processo mentale, ne consegue che:

> la percezione di alto livello è caratterizzata dal fatto di essere di tipo semantico: essa implica il fatto di estrarre il *significato* delle situazioni. Quanto maggiore è l'elaborazione semantica, tanto maggiore è il ruolo che vi rappresentano i *concetti*, e quindi la portata delle influenze top-down. La comprensione delle situazioni nel loro insieme rappresenta il genere di percezione più astratto possibile, e anche il più flessibile. (*Ivi*, p. 190)

La prima caratteristica dei modelli cognitivi che tentano di riprodurre questa capacità umana è, dunque, quella di fornire una tentativo di spiegazione e di messa alla prova, attraverso la simulazione, di alcune teorie del significato. Inoltre, un altro aspetto centrale consiste nel mostrare non solo come la conoscenza possa essere implementata in un programma, ma anche che ruolo *attivo* essa svolga nei processi di pensiero. La questione dell'analogia riveste una posizione subordinata a tali obiettivi. Essa, in tale prospettiva, è funzionale allo studio dei processi in oggetto. D'altra parte, l'analogia gode di questa caratteristica perché la percezione di alto livello è vista come analogia *lato sensu*, nel senso *del* processo che produce l'analogia, e, viceversa, l'analogia, nel senso del fare analogie, è considerata *il* nucleo essenziale dei processi cognitivi di alto livello[6]. Cerchiamo di chiarire la questione con un esempio.

Immaginiamo di ascoltare la seguente affermazione: "Stalingrado è stata la Caporetto di Hitler". Per capire questa espressione è necessaria un'ampia dose di conoscenze contestuali. Innanzitutto, ci occorre avere un certo numero di competenze relative alla comprensione della lingua in cui viene pronunciata, in questo caso l'italiano, e alla struttura sintattica della frase: un soggetto, una copula, un predicato nominale e un complemento di specificazione. Dato per scontato che questi due tipi di conoscenze siano in nostro possesso, ci occorre ancora un bagaglio di significati per arrivare a capire l'espressione, ovvero una conoscenza semantica che ci permetta di comprendere a quali eventi si riferisce l'affermazione. A questo punto saremmo tentati di pensare che non ci serve altro per capire l'affermazione. In realtà, è necessario ancora un passo ulteriore, che ci porti a comprendere la natura della relazione in cui sono stati posti i concetti. Questo passo è la costruzione di una struttura analogica che ponga in evidenza gli aspetti di somiglianza, di *mappatura concettuale*, individuabili fra le molte differenze che intercorrono fra i due accadimenti. Infatti, pur se entrambe sono state celebri battaglie, l'una della seconda guerra mondiale, l'altra della prima, non sono molti i punti che hanno in comune, ma sono proprio questi a costituire l'aspetto della situazione posto in evidenza dall'affermazione.

I due accadimenti riguardano guerre e tempi diversi, sono combattuti da eserciti e nazioni diverse, non condurranno gli eserciti sconfitti ad uno stesso esito nel lungo periodo, né hanno lo stesso peso sul conflitto, inteso in senso globale, in cui sono avvenuti e molte altre differenze possono essere trovate. Eppure esistono alcuni aspetti per cui possono essere accostati. Si tratta appunto di due sconfitte di eserciti che stavano avanzando in territorio nemico, sono due disfatte che implicano un immediato abbandono delle posizioni raggiunte, costringono gli eserciti alla ritirata nella stessa direzione (da est verso ovest) per un buon numero di chilometri, sono causa di un numero elevato di vittime.

[6] Idea portante del saggio di Hofstadter *Analogy as the core of cognition*, contenuto in Holyoak, Gentner, Kokinov (2001) e ridiscussa recentemente (gennaio 2006) in una delle President Lecture di Stanford.

Il processo di mappatura concettuale ha una doppia funzione. In prima istanza crea una serie di relazioni biunivoche fra elementi diversi delle due situazioni. Allo stesso tempo, nel fare questo fa anche risaltare le differenze fra i due eventi considerati. La creazione di analogie, perciò, può essere considerato come un processo di polarizzazione somiglianze/differenze, che si produce a seguito della costruzione di una corrispondenza diretta e biunivoca fra due domini distinti per una qualche dimensione del tempo e/o dello spazio (nulla esclude, dunque, che si possa parlare anche della stessa situazione in due momenti di tempo diversi)[7]. Tale polarizzazione è il risultato di un processo costruttivo di rappresentazioni adeguate alla mappatura, cioè all'istituzione della relazione di corrispondenza. La "costruzione di rappresentazioni adeguate alla mappatura" è ciò in cui consiste propriamente la percezione di alto livello.

Chalmers, French e Hofstadter distinguono a proposito due parti essenziali del procedimento analogico. La prima è «il processo di *percezione di una situazione*, che consiste nel considerare i dati relativi a una data situazione, quindi filtrarli e organizzarli in vari modi per arrivare a una rappresentazione appropriata al contesto particolare». C'è, poi, «il processo di *proiezione per mappe*, che consiste nel trovare le corrispondenze appropriate tra gli elementi dell'una e quelli dell'altra, creando così l'accoppiamento che chiamiamo analogia». I due processi non sono separabili dal punto di vista operativo, anzi «sembrano avere interazioni profonde» (*ivi*, p. 199). Tale inseparabilità deve riflettersi nel modello simulativo. Costitutivamente essa ha due direzioni. Il secondo processo dipende in maniera significativa dal primo il quale consiste, a conti fatti, nella costruzione di una rappresentazione adeguata e utilizzabile per il processo di mappatura. Fare un'analogia dipende in senso stretto dalla percezione di alto livello. Tuttavia, poiché i due processi tendono in questa impostazione a essere considerati lo stesso da un punto di vista più generale, è possibile affermare che la percezione di alto livello dipende strettamente dall'attività del fare analogie. Tra le due operazioni si crea, pertanto, una specie di circolo autoreferenziale di rafforzamento reciproco, raffigurabile attraverso un ciclo virtualmente interminabile (fig. 2.1).

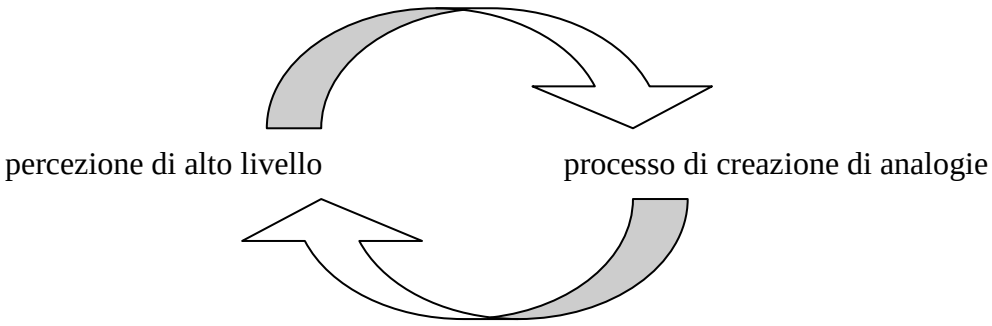

Fig. 2.1

[7] Secondo tale prospettiva l'intero procedimento ricorda da vicino l'attività di *frame blending*, cioè di mescolanza di strutture, discussa da Fauconnier e Turner (2002). La mappatura concettuale, infatti, ne costituisce l'operazione fondamentale.

Maggiore chiarezza sul rapporto che intercorre fra i due termini del processo circolare è ottenibile una volta esplicitato il ruolo che vi assumono i concetti. Si è detto che il loro intervenire segna il punto in cui la percezione di basso livello si trasforma in percezione di alto livello. In che modo, tuttavia, può essere caratterizzata la loro funzione? La mera presenza di un concetto in un compito conoscitivo non produce alcunché. Esso deve necessariamente ricoprire una funzione "attiva" per il prodursi della conoscenza che costituisce il risultato della percezione di alto livello, e questo sia per quanto riguarda l'aspetto della categorizzazione, sia in relazione alla proiezione di strutture fra una situazione e l'altra. La nozione chiave è quella di "slittamento concettuale", che si caratterizza come: «*la rimozione di un concetto indotta dal contesto ed effettuata da un altro concetto strettamente connesso al primo*, all'interno della rappresentazione mentale di una situazione» (Hofstadter & FARG, p. 216).

Nella mappatura è ben evidente *come* si realizzi, cioè *a che cosa* si applichi questa operazione. Tuttavia, sembra non del tutto chiaro il modo in cui essa abbia a che fare con la percezione di alto livello, se non per il fatto che, per essere tale, questa deve coinvolgere i concetti. La nozione di "slittamento concettuale", che come si vedrà rappresenta una svolta[8] nella costruzione dei modelli (sub)cognitivi del fare analogie, viene chiarita ulteriormente da Mitchell, la quale rivendica «l'ubiquità e la centralità della percezione di alto livello e dello slittamento concettuale in tutti gli aspetti del pensiero, dagli atti basilari e ordinari di riconoscimento e categorizzazione alle caratteristiche elusive e apparentemente mistiche dell'*insight* e della creatività» (Mitchell, 1993, p. 2). Infatti, lo slittamento concettuale e la percezione di alto livello costituiscono il nucleo del pensiero e convergono nel processo di produzione di analogie: «poiché il fare analogie consiste interamente nel percepire somiglianze tra cose che sono differenti, un'analogia impone una certa pressione ai concetti affinché slittino in concetti correlati» (*ivi*, p. 5).

L'importanza del contesto appare un fatto ineliminabile del fare analogie. Senza di esso non si avrebbe alcuno slittamento concettuale e i concetti svolgerebbero un ruolo solo nell'essere attivati in base al ritrovamento di loro istanze nell'ambiente percettivo. La presenza di un contesto costituito da una rete semi-variabile e relazionale di concetti garantisce che lo slittamento sia possibile e l'operazione di mappatura sia compiuta. Tale elemento, perciò, sarà uno dei tratti fondamentali dei modelli che vogliono simulare il meccanismo di produzione delle analogie.

Tuttavia, per delineare in maniera più precisa la nozione di percezione di alto livello occorre esplicitare i diversi gradi dello spettro dal concreto all'astratto che esprime il suo ambito di applicazione. French ne propone una classificazione in nove tipi (French, 1995, pp. 11-13):

- **riconoscimento**: è il processo per cui un'entità è riconosciuta appartenere a una determinata categoria *senza che la categoria risulti modificata*. Avviene quando si percepisce un cane o

[8] In particolare dalla progettazione e implementazione di COPYCAT, il primo modello di *analogy-making* che condivide questa impostazione.

un albero molto vicini al concetto prototipico posseduto (e questo anche se il cane è disegnato o l'albero compare in fotografia);

- **generalizzazione**: è il processo per cui un'entità è riconosciuta appartenere a una determinata categoria *con l'apporto di alcune modifiche alla categoria*. Si consideri, ad esempio, il caso di un individuo che vede per la prima volta una mangrovia e la classifichi come albero poiché possiede molte qualità in comune con questo concetto, al quale aggiunge la caratteristica *modificante* di mettere le radici in acqua salata;
- **somiglianza superficiale**: è il caso in cui due situazioni vengono considerate analoghe solo in base a caratteristiche superficiali condivise e *non* a quelle più profonde e strutturali. Ne sono esempi le similitudini utilizzate in poesia ("la luna è una bianca fetta di formaggio gruviera") o gli epiteti conferiti, scherzosamente o meno, agli individui nei contesti sociali e basati sulle somiglianze fisiche ("è una balena!");
- **pluralizzazione**: è il processo per cui la categoria riferita a una singola situazione, ad esempio il nome di una persona o di una personaggio, viene utilizzata per riferirsi a un insieme di persone *sulla base di una caratteristica ben definita e conosciuta* ("sei sempre il Grillo Parlante della situazione!") e senza la necessità che le altre caratteristiche siano rilevanti o addirittura note;
- **analogie "anch'io"**: riguardano le situazioni in cui si afferma la propria intenzione di fare qualcosa rendendola analoga all'intenzione espressa da qualcun altro di fare qualcosa di simile. Ad esempio, fuori da un supermercato due amici si incontrano e, dopo aver conversato per un po', uno dice: "Vado a prendere la mia macchina" e l'altro risponde: "Anch'io", intendendo, ovviamente, l'intenzione di prendere la propria macchina e non quella dell'amico[9];
- **supertraslazioni**: in esse, simili alle pluralizzazioni, un concetto gioca il ruolo di un altro in un determinato contesto. Schematicamente: B è l'A di Y, cioè B fa la parte di A nel contesto Y. Il contesto in cui A gioca il suo ruolo rimane generalmente implicito. Nell'affermazione: "Napoleone è l'Alessandro Magno dell'Europa moderna" il contesto in cui visse e agì Alessandro Magno rimane sullo sfondo e si suppone che sia implicitamente riconosciuto e condiviso;
- **analogie caricaturali**: sono analogie create appositamente per mettere in luce a fini esplicativi alcune caratteristiche implausibili di una situazione complessa. La realtà controfattuale che costruiscono e che viene opposta all'affermazione dell'interlocutore gioca sul contrasto ironico con la situazione di partenza e si avvale di stereotipi. Si consideri il caso di qualcuno che ci dica: "Paolo Rossi è il giocatore di calcio italiano più noto di tutti i tempi", a cui potremmo rispondere: "Suvvia! È come dire che la Maserati è la macchina italiana più conosciuta al mondo";

[9] Una rassegna di questo tipo di analogie si può trovare in Hofstadter (1991).

- **analogie esplicative**: come le precedenti sono analogie create per spiegare una situazione di difficile comprensione, senza, però, alcun intento di forzatura ad accettare l'accostamento, bensì con l'utilizzo di fatti e conoscenze relative alla propria esperienza personale. Ad esempio, nell'affermazione: "la diffusione di internet alla fine del ventesimo secolo è stata come l'introduzione dei caratteri a stampa nel quindicesimo secolo" la corrispondenza viene usata per spiegare gli effetti e le implicazioni di questa complessa trasformazione tecnologica e sociale;
- **rievocazioni episodiche**[10]: il processo per cui alcune caratteristiche della situazione presente ci ricordano una situazione passata della nostra esperienza personale.

Come è facile vedere, tale casistica esplicita un senso molto lato di analogia, che va dalla ascrizione di un input esterno ad una determinata categoria fino all'accostamento di un'esperienza presente con una passata. Il minimo comune denominatore di questo processo risiede nel fatto che ogni punto dell'elenco riguarda una relazione che si instaura fra due insiemi attraverso la messa in relazione di strutture, sia quelle costituite di elementi percettivi con un concetto specifico, sia quelle concettuali più complesse, costruite o meno in maniera consapevole, nel processo analogico. Sono i concetti, e perciò la conoscenza che un individuo possiede, a guidare in ogni caso tale processo, anche nel caso dei livelli più bassi, come nel riconoscimento, che deve intendersi alla stregua di una riconduzione di un insieme variabile di tratti a una struttura invariante, un concetto, i tratti del quale siano costitutivi, anche se non in maniera esuastiva e determinata una volta per tutte, bensì variabile a seconda del contesto della sua possibile definizione.

French suggerisce che sia la quantità di slittamento concettuale coinvolto nell'analogia a differenziare i diversi casi di analogia proposti e individua tre tipi di slittamento (*ivi*, pp. 3-5): esportazione, trasporto, importazione. Il primo tipo è un processo di astrazione dalla situazione concreta a uno schema astratto. Nello schema i nomi vengono rimpiazzati da variabili, per cui il processo è definito complessivamente: "astrazione e variabilizzazione". Per tornare all'esempio precedente, si può rispondere alla domanda: "quale è la Caporetto di Hitler?", procedendo in questo modo. Si consideri la situazione iniziale descritta dalla locuzione:

i) la Caporetto dei Savoia

la quale può essere trasformata nel seguente schema concettuale (tenendo sempre presente che la trasformazione è un'esplicitazione di tratti non univoca)

[10] Le rievocazioni episodiche sono trattate diffusamente in Schank (1982).

ii) la grande sconfitta militare, che produce una ritirata da est a ovest in cui muoiono moltissimi soldati e che porta a un arretramento dei confini sui territori conquistati dopo un lungo periodo di avanzamento vittorioso, di X

in cui “Caporetto” viene sostituita da “grande sconfitta militare, ecc.” (procedimento astrattivo) e “i Savoia” da X (procedimento di variabilizzazione).

Lo slittamento di trasporto è nient’altro che il processo per cui la variabile X viene rimpiazzata e nuovamente vincolata con una costante:

iii) la grande sconfitta militare di Hitler.

che, come si vede, non deve necessariamente rispettare tutti i vincoli grammaticali (Hitler è un individuo singolo, i Savoia un nome collettivo di dinastia), ma è vincolata a ricoprire lo stesso ruolo nella nuova situazione, quello di capo supremo dell’esercito in guerra. A questo punto, l’applicazione di uno slittamento di importazione rende possibile il completamento dell’analogia, ed è qui che risiede il nucleo rilevante del processo analogico. Infatti, a “la grande sconfitta militare, ecc.”, concetto che rappresenta l’astrazione di Caporetto, va sostituito il concetto di un evento specifico che stia in una qualche relazione con Hitler. A questo punto una serie di pressioni spingono a considerare Stalingrado come l’equivalente di Caporetto nella situazione denotata da “Hitler”, ovvero “comandante in capo di un esercito invasore con desideri di conquista verso oriente all’interno di territori nemici”. Questo avviene nonostante si presentino molti elementi di divergenza fra le due situazioni, le quali sono messe in evidenza, unitamente alle somiglianze, dal processo di elaborazione in cui consiste lo slittamento di importazione.

French rappresenta le fasi del processo nel modo indicato in figura 2.2 (adattata da French, 1995, p. 4):

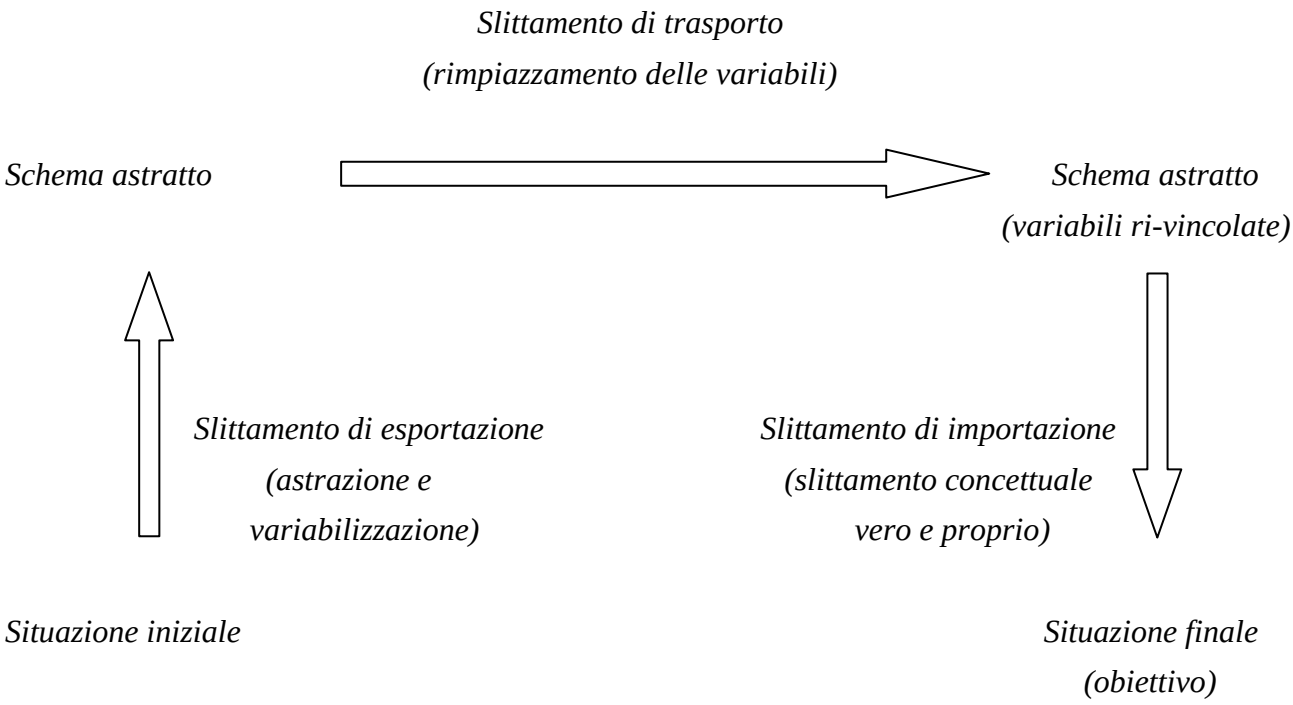

Fig. 2.2

La figura sottolinea il fatto che lo slittamento concettuale, ovvero l'attivazione di un nuovo concetto che abbia *lo stesso ruolo* ricoperto da quello nella situazione iniziale una volta istanziate nuovamente le variabili, avviene solo nell'ultima fase, quella dell'importazione. Tutto ciò sembra suggerire l'idea che l'analogia si dia solo per i tipi più complessi di percezione di alto livello, cioè quelli che comprendono processi di mappatura costruiti in maniera più o meno esplicita. In realtà, anche per i più basilari processi di riconoscimento e categorizzazione si può parlare di analogia, come sembra suggerire Hofstadter (Hofstadter,1981) quando afferma che il nostro utilizzare una stessa parola, ad esempio "gatto", per riferirci a tutti i possibili gatti che ci capita di incontrare, non è altro che mettere in atto un'analogia fra l'input percepito al momento presente e i ricordi di input simili immagazzinati in memoria ed etichettati con la medesima parola, che nomina il medesimo concetto astratto per tutti i gatti possibili. Non esistendo, infatti, due gatti perfettamente identici, siamo costretti a ricorrere a un'analogia fra l'informazione percettiva e quella memorizzata al momento in cui il processo di riconoscimento viene compiuto. Questa operazione si complica via via che risaliamo la scala di complessità dei tipi di analogia (e di percezione di alto livello). Nel processo di categorizzazione, ad esempio, l'analogia è più complicata, perché il processo di confronto fra input percettivo e informazione contenuta nella memoria modifica quest'ultima aggiungendo delle caratteristiche al concetto[11].

[11] Lasciamo per ora da parte le questioni molto dibattute e ancora aperte sulla natura dei concetti proposte da teorie psicologiche che oscillano, tanto per fare solo un esempio, fra una spiegazione che pende verso una prototipicità dei concetti immagazzinati in memoria e un'altra che li vede come collezioni di esempi ricavate dall'esperienza passata.

Il fatto che tali processi siano in larga parte elusivi del ragionamento cosciente testimonia dell'assenza di consapevolezza nel compierli. Essi appaiono tanto più irriflessi quanto più immediati. Tuttavia, la natura prevalentemente inconscia del processo di creazione di analogie, inteso in senso lato, non è una questione di complessità. In realtà, ciò che accomuna i vari aspetti di questa teoria dell'analogia come procedimento costitutivo ed essenziale del pensiero non è solo il suo permeare qualsiasi atto di percezione di alto livello (ovvero concettuale a un qualche grado più o meno elevato di complessità), ma anche il fatto che, in ogni caso, i processi che conducono alla costruzione di analogie si basano su micro-operazioni cognitive, anche nel caso in cui la costruzione sembri essere effettuata del tutto consciamente. In ciò si può riscontrare uno dei principi teorici basilari dell'approccio subcognitivo alla spiegazione dei meccanismi della pensiero.

In seguito verranno presi in esame un serie di modelli cognitivi che hanno lo scopo di mostrare come tale teoria possa venire implementata sul calcolatore al fine di riprodurre questa caratteristica, sfuggente in termini consapevolezza, del pensiero umano. Rimane ora da chiarire in che modo va intesa la nozione di "somiglianza" che sta alla base di, e rende possibile il, processo di creazione di analogie. Tale nozione si rivela, infatti, problematica da diversi punti di vista. Goodman, ad esempio, ha argomentato ampiamente contro varie accezioni del concetto di somiglianza (Goodman, 1972). Egli asserisce, fra le altre cose, che non è la somiglianza che spiega la natura iconica di un simbolo o quella realistica di una immagine, bensì al contrario, è tale relazione a fondarsi su queste. D'altra parte, non è la somiglianza il criterio per cui due istanze possono essere riferite allo stesso tipo ideale o rendere due comportamenti o esperimenti scientifici lo stesso comportamento o lo stesso esperimento effettuato in momenti diversi; né essa è alla base di una generalizzazione induttiva, piuttosto quest'ultima può essere considerata uno dei criteri per l'attribuzione di somiglianza fra situazioni diverse.

Generalizzando, non è possibile affermare che due entità o due situazioni sono simili soltanto in base al fatto che «posseggono caratteristiche comuni» (*ivi*, p. 443). È sempre possibile, infatti, trovare una medesima proprietà condivisa da due oggetti. Ad esempio, se prendo una sveglia e una penna, posso dire che sono simili perché entrambe sono manufatti artificiali, o, anche, che sono entrambi oggetti singoli, ovvero che appartengono entrambe all'insieme formato dagli insiemi con un solo elemento. Pertanto, appare evidente come su base estensionale sia sempre possibile rendere qualcosa simile a qualcos'altro. Da ciò deriva che per stabilire una relazione di somiglianza occorre riferirsi non soltanto alle proprietà intensionali di un concetto, ma alle «proprietà *importanti* – o meglio, considerando non la somma ma l'importanza globale delle proprietà condivise» (*ivi*, p. 444), siano esse estensionali o intensionali. Questo equivale a dire che due entità o situazioni possono essere messe in relazione di somiglianza sulla base di *criteri rilevanti che di volta in volta*

Entrambe queste teorie appaiono, in ogni caso, compatibili con questo approccio al problema della percezione di alto livello.

dipendono dal contesto (percettivo di basso e alto livello o anche relativo agli scopi e agli interessi di chi ravvisa o cerca di stabilire una relazione di somiglianza).

In tale prospettiva il fare analogie diviene inevitabilmente il meccanismo base di ogni processo di pensiero, in quanto l'analogia può essere intesa come l'espressione del contesto in base a cui istituire un relazione di somiglianza, e, allo stesso tempo ma in direzione opposta, ogni percezione di alto livello di qualunque tipo non può avvenire al di fuori di un contesto, implicito od esplicito che sia. Pertanto, se il ragionamento per analogia è un ragionamento per somiglianza esso è imprescindibile dal contesto in cui viene attuato, il quale d'altra parte definisce i limiti *entro cui* può essere attuata l'operazione di astrazione, che prelude allo slittamento concettuale e che si configura come una sorta di *parziale affrancamento* dal contesto stesso. Solo attraverso questa "parzialità" è possibile creare strutture che siano sufficientemente ampie da poter racchiude diversi contenuti semantici e allo stesso tempo non così tanto da rendere indifferente, perché non vincolata, la relazione di somiglianza. Tale operazione conduce in definitiva al riconoscimento di *ruoli* all'interno di *strutture*. Così si esprime Hofstadter al riguardo:

> [...] sembra che quasi ovunque, dentro alle rappresentazioni interne dei concetti, esistano sottostrutture che sono *relativamente* indipendenti dalle strutture di cui fanno parte. Una sottostruttura di questo tipo è modulare, cioè è esportabile dal contesto originale ad altri contesti. Una tale sottostruttura può essere considerata in modo autonomo e conveniamo di chiamarla "ruolo". Un ruolo, dunque, è una "descrizione modulare" naturale dotata di una grande mobilità, potendosi spostare facilmente dal contesto in cui è nata ad altri contesti che a prima vista non si sospetterebbero. (Hofstadter, 1981, p. 140 [corsivo mio])

La ricerca di una struttura adatta alla mappatura analogica si basa, pertanto, sulla individuazione di tali ruoli. Essi si caratterizzano come sottostrutture dei concetti e sono il risultato dell'operazione di astrazione (slittamento di esportazione) vista in precedenza. Di conseguenza, è possibile affermare che, in primo luogo, se è pur vero che la mappatura è la vera essenza dell'operazione di creazione di analogie, appare evidente che la costruzione di strutture, la percezione di uno schema astratto, riveste il ruolo di condizione necessaria al processo. In secondo luogo, la costruzione di tali strutture appare guidata sia dall'analogia, o, meglio, dal contesto analogico, nel senso proporzionale di rapporto che deve essere trasferito da due entità o situazioni ad altre due entità o situazioni, sia, e soprattutto, dal contesto epistemico in cui l'analogia viene compiuta. La percezione di strutture astratte, infatti, si configura come estrapolazione dei ruoli salienti all'interno del contesto epistemico (del sistema cognitivo che la attua), il cui apporto garantisce non solo la possibilità che i ruoli vengano esplicitati, ma anche che tali ruoli siano quelli più adatti, nel senso di rilevanti *rispetto a un qualche criterio*, per la costruzione dell'analogia.

Per tali ragioni, i modelli che implementano l'approccio subcognitivo sono solo in parte (anche se per una buona parte) modelli del fare analogie, ma sono *tutti* modelli della percezione di alto livello.

Il loro dominio di applicazione è in genere ristretto e viene definito "microdominio". Prima, però, di passare alla discussione di questo loro secondo aspetto caratteristico, è utile considerare quali altri tipi di modelli computazionali dell'analogia sono stati proposti nel corso della ricerca in IA, per marcarne somiglianze e differenze con quelli oggetto della presente trattazione.

2.3 L'intelligenza artificiale e il ragionamento analogico

Una lunga tradizione di studi sul ragionamento analogico si sviluppa lungo tutto il corso dell'IA, in misura anche superiore rispetto alle ricerche sullo stesso tema effettuate in ambito strettamente psicologico[12]. Intuitivamente, l'idea che un programma possa apprendere, e quindi immagazzinare in una qualche forma di memoria, soluzioni, metodi o rappresentazioni per poi riutilizzarli nelle successive applicazioni, pone il problema di come collegare questa conoscenza conservata con la disponibilità immediata in riferimento alla situazione corrente di esecuzione del programma. Il richiamo per via analogica della conoscenza posseduta appare una delle soluzioni possibili e la questione di come ciò possa venire implementato dà l'avvio ad un filone di ricerca interamente dedicato al tentativo di progettazione e realizzazione di modelli computazionali di ragionamento analogico. Lo stesso Minsky, uno dei padri fondatori dell'IA, così si esprimeva commentando ANALOGY, il primo programma in grado di svolgere compiti che coinvolgono procedure analogiche:

> [...] sta diventando chiaro che il ragionamento analogico stesso può essere uno strumento importante ai fini dell'allargamento dell'intelligenza artificiale. Ritengo che sarà finalmente possibile per i programmi, per mezzo del ricorso al ragionamento analogico, l'applicazione dell'esperienza che hanno acquisito attraverso la soluzione di un tipo di problema alla soluzione di problemi del tutto differenti. (Minsky, 1966, p. 251)

Per l'analogia viene, dunque, rivendicato un ruolo centrale nell'apprendimento. Tuttavia, non è soltanto il *machine learning* a occuparsi di ragionamento analogico. Il risultato di un'analogia può essere considerato come ciò che un programma apprende, e conserva, in merito a una certa situazione o a un certo metodo di soluzione di un problema, ma anche come il presupposto per nuove analogie in altri domini. Ne consegue che è possibile valutare i modelli di implementazione del procedimento analogico in virtù della loro adeguatezza nello sviluppare le diverse operazioni in

[12] Questo potrebbe far pensare a una maggiore considerazione "filosofica" del problema dell'analogia e potrebbe essere considerata una prova a favore del fatto che, a dispetto delle dichiarazioni in proposito, l'IA ha avuto e ha ancora a che fare molto più con problemi filosofici che non psicologici. Per una trattazione storico-teoretico-filosofica del problema dell'analogia si rinvia al densissimo libro di Melandri *La linea e il circolo* (recentemente riedito. Si veda Melandri, 2004).

base a cui il processo di produzione di analogie si compie. Hall ne individua quattro (Hall, 1989, p. 43):

- **riconoscimento**, in base ad opportuni parametri, di una situazione nota come *adeguatamente analoga* ad una situazione obiettivo non conosciuta;
- **elaborazione** di una proiezione per mappe fra la situazione nota e quella da analizzare;
- **valutazione** della mappatura nel contesto d'uso dell'analogia, che culmina con l'espressione di un giudizio e determina eventuali operazioni di modifica ed estensione della proiezione;
- **consolidamento** dell'esito dell'analogia, in termini sia di adeguatezza del risultato sia delle strutture relazionali astratte prodotte.

L'analisi comparativa di Hall si riferisce a modelli computazionali simbolici, ma si adatta anche alla valutazione di modelli di tipo connessionista o ibridi. Kokinov e French suddividono il processo di costruzione di analogie in sei sottoprocessi nella loro disamina di modelli computazionali che si avvalgono di approcci differenti (Kokinov, French, 2003, pp. 114-115):

1. costruzione di rappresentazioni;
2. recupero;
3. mappatura;
4. trasferimento;
5. valutazione;
6. apprendimento.

Come è facile vedere, i primi due sono una suddivisione del processo di riconoscimento di Hall in due sottofasi, mentre la mappatura e il trasferimento specificano due diversi momenti del processo di elaborazione. Kokinov e French sottolineano come al processo di costruzione di rappresentazioni, il quale corrisponde a quello visto in precedenza di percezione di strutture astratte, non è stata in realtà prestata molta attenzione nel corso degli anni a dispetto dell'elevato numero di modelli proposti per il ragionamento analogico. Inoltre, essi fanno notare che in molti casi la procedura di recupero, diversamente da quella di mappatura, è basata sulla somiglianza superficiale fra una situazione nota, conservata in memoria, e quella meno nota oggetto d'analisi da parte del programma, ma avanzano anche l'ipotesi che la costruzione di una struttura relazionale di corrispondenze fra elementi di due situazioni diverse è un procedimento composito basato in ugual misura sia su elementi superficiali, sia su isomorfismi strutturali[13], sia sull'importanza dal punto di vista pragmatico degli elementi presenti nell'obiettivo. Il trasferimento, infine, è il vero processo di

[13] L'importanza degli elementi strutturali "profondi" dal punto di vista cognitivo nella costruzione di analogie fra entità e situazioni differenti è stata sostenuta soprattutto dalla Gentner (1983).

incorporamento di nuova conoscenza nel, e rispetto al, dominio dell'obiettivo, sulla base di ciò che è rilevante nella situazione di partenza e che si ipotizza avere una controparte nella situazione *target*. È questo il sottoprocesso che rende effettivamente possibile l'allargamento della conoscenza.

Al di là delle caratteristiche specifiche che ogni modello presenta, o che possiede in misura maggiore o minore, è possibile raggruppare i modelli computazionali del ragionamento analogico secondo una tipologia ormai familiare all'interno dell'IA: i modelli simbolici, quelli connessionisti e quelli ibridi. Consideriamone alcuni esempi per ogni tipo.

2.3.1 Modelli simbolici

A questo gruppo appartiene ANALOGY di Thomas Evans, uno dei primi programmi di implementazione del ragionamento analogico (Evans, 1968). Il suo dominio di azione era quello della geometria. Al programma venivano sottoposti quesiti in cui a tre figure note, A, B e C veniva chiesto di associarne una quarta, D, in modo che il rapporto fra C e D fosse lo stesso rintracciabile fra A e B. ANALOGY aveva a disposizione cinque possibili risposte e procedeva al rinvenimento della soluzione migliore fra quelle proposte, costruendo una rappresentazione delle relazioni fra le figure geometriche all'interno delle figure complessive A, B, C e delle cinque soluzioni proposte. La descrizione di ogni figura complessiva era data da relazione diadiche formulate nel calcolo dei predicati (del tipo RIGTH(x,y) o INSIDE(w,z)) e il rapporto intercorrente fra A e B era espresso da una regola (un condizionale) in cui l'antecedente era costituito dai predicati rappresentanti la descrizione di A e il conseguente da quelli che rappresentavano la descrizione di B. A questo punto il processo di mappatura procedeva a un confronto su un doppio livello: quello degli oggetti e quello del ruolo degli oggetti nelle relazioni. In tal modo, veniva creata una relazione di corrispondenza fra gli elementi di A e C e il programma procedeva alla valutazione di quale figura, in base alla regola che esprime il rapporto fra A e B, era più adatta a occupare il posto di D nell'analogia. Il doppio livello di comparazione consentiva di muoversi fra somiglianze oggettuali e somiglianze relazionali, queste ultime prese in considerazione prima delle altre nel valutare in che modo A e C rivestivano il medesimo ruolo come "situazioni di partenza". Una volta trovata la corrispondenza fra le relazioni all'interno delle figure complessive ANALOGY procedeva al riempimento delle relazioni con gli oggetti e stabiliva collegamenti fra gli oggetti di A e C. Infine, in base all'assunto che oggetti dello stesso tipo, cioè con *lo stesso ruolo*, devono occupare gli stessi posti nelle relazioni dei due rapporti A:B e C:D, il programma selezionava una risposta fra quelle possibili.

Come è evidente, anche in un semplice compito come quello descritto, che pure non prevede la costruzione di una situazione (figura complessiva) finale, ma solo la sua scelta all'interno di un insieme molto limitato, l'operazione di analogia richiede una molteplicità di procedure operative

intrecciate. ANALOGY sfrutta un processo *bottom up* di costruzione della rappresentazione della situazione data, in cui si passa dal rinvenimento di semplici oggetti e relazioni fra oggetti alla formulazione di regole, ottenendo come risultato l'individuazione del ruolo dell'oggetto all'interno della sua figura complessiva. Tale intuizione operativa non è stata ripresa da altri modelli di tipo simbolico classico. In quasi tutti gli altri programmi progettati e implementati nel corso di trenta anni[14], i quali si muovono in domini diversi e sfruttano tecniche diverse di elaborazione, la rappresentazione della situazione è consegnata al programma, che in genere parte dal processo di recupero (*retrieval*) di vecchie situazioni per poi metterle a confronto con quelle nuove. Le vecchie situazioni sono date in forma di collezioni di proposizioni o di reti proposizionali corredate di vincoli, preformate e adattate allo scopo che si intende conseguire.

In termini generali, il processo di elaborazione consiste nel confrontare lo schema proposizionale in cui è espressa la situazione obiettivo e quello in cui è espressa la situazione recuperata per l'analogia. Questo avviene ad esempio nel modello di Winston (1982, 1986), in cui il ragionamento analogico è utilizzato per implementare procedure di apprendimento automatico e che si avvale di un approccio *bottom up* per il recupero di situazioni note; o nel programma CARL di Burnstein (1986), che impiega analogie fra situazioni strutturate in *frame* per ricavarne concetti (astrazioni concettuali). Un approccio del medesimo tipo è stato utilizzato anche nel campo della deduzione automatica, come, ad esempio, in NLAG di Greiner (1988), che sfrutta analogie tra fatti espresse in forma proposizionale per guidare il processo deduttivo; o in quello del *problem solving*, come in ANA di McDermott[15], che utilizza un sistema a regole di produzione che codificano la conoscenza del programma e vengono impiegate quando le loro condizioni sono soddisfatte da (cioè combaciano con) elementi presenti nella memoria di lavoro; o anche per programmi che ricadono nell'orbita del ragionamento secondo casi (*Case-Based Reasoning*, più noto come CBS), come in MEDIATOR (Kolodner, Simpson, Sycara-Cyranski, 1985), in cui gli episodi immagazzinati nella memoria a lungo termine vengono confrontati con la situazione da analizzare attraverso una mappatura fra strutture *frame* (*matching* degli *slot*); o, infine, in quella del *planning*, come nei metodi di risoluzione suggeriti da Carbonell (1983).

Al di là delle differenze algoritmiche per quanto riguarda l'elaborazione del processo analogico, tutti questi modelli condividono l'idea di una mappatura fra situazioni diverse appartenenti a domini diversi. Tale mappatura avviene fra sistemi proposizionali o a strutture *frame* già date e si realizza in una corrispondenza biunivoca oggetto-oggetto e relazione-relazione; oppure, al contrario, si realizza nella costruzione di due strutture astratte – sempre nella forma del calcolo dei predicati – a partire dalla situazione nota e da quella obiettivo (dalla situazione nota A è derivata A' e dalla situazione obiettivo B è derivata B') e si procede alla messa a confronto delle strutture costruite con

[14] Si rimanda a Hall (1989) per una rassegna e una discussione comparativa dei modelli computazionali del ragionamento analogico sviluppati in questo periodo.
[15] Descritto in Hall (1989).

quelle già date (A' con B e B' con A) per misurarne l'eventuale discrepanza, al fine di ridurla con il procedimenti come quello dell'analisi mezzi-fini[16].

A questa impostazione appartengono anche il modello computazionale più noto del ragionamento analogico, Structure Mapping Engine (SME) e le sue estensioni, basati sulle teorie della Gentner in merito alla proiezione per mappe di strutture (Gentner, 1983). L'idea di base di questa teoria è che nel processo analogico abbia predominanza la componente strutturale profonda delle situazioni messe in relazione piuttosto che i loro aspetti superficiali. In altri termini, non sono le proprietà superficiali di un oggetto a essere rilevanti per la costruzione di un "ponte" fra due situazioni, ma le reciproche connessioni fra oggetti all'interno di ogni situazione. Per tale ragione, viene data priorità alle corrispondenze fra relazioni, piuttosto che a quelle fra proprietà. Ancora, l'operazione di mappatura delle relazioni di ordine superiore, i cui argomenti sono a loro volta relazioni, è anteposta, attraverso la messa in corrispondenza di queste ultime fra dominio iniziale e dominio obiettivo, a quella delle relazioni di livello inferiore. Inoltre, i vari passi implementativi sono tra loro isolati e basati su meccanismi indipendenti. Recupero, proiezione per mappe e valutazione dell'analogia costruita avvengono separatamente e sequenzialmente, conferendo una certa rigidità al processo.

Il modello SME, almeno nella sua prima versione (Falkenhainer, Forbus, Gentner, 1989), si avvale, dunque, di un tipo di elaborazione *top down*, in cui le strutture più generali e più astratte (relazioni di relazioni) hanno un ruolo guida per il processo di mappatura, nel quale esclusivamente consiste il procedimento di costruzione analogica. Inoltre, almeno a livello più elevato, solo relazioni identiche nei due domini vengono poste in corrispondenza, concedendo ben poco spazio di manovra, e perciò poca variabilità e poca creatività, al programma, estromettendo quelle che sono le caratteristiche peculiari del ragionamento analogico. Così, nel caso di una delle più famose prestazioni riuscite del programma, la costruzione di un'analogia fra il sistema solare e il modello dell'atomo di Rutheford, SME non fa altro che mettere in corrispondenza relazioni del tipo:

i) Causa (Gravità, Attrae (Sole, Pianeta))

con altre, identiche dal punto di vista sintattico-formale, ma soddisfatte da argomenti diversi e appartenenti all'altro dominio, come ad esempio:

ii) Causa (Segno opposto, Attrae (Nucleo, elettrone)

fornite in maniera preconfezionata dall'utente o dal programmatore. Il processo, che pure consiste nella costruzione di diversi sistemi di corrispondenza di relazioni e nella valutazione del migliore, si riduce alla sovrapposizione di relazioni identiche, senza che il programma abbia per nulla

[16] Un modello di questo tipo è JCM (Becker, 1973).

sviluppato una benché minima *comprensione semantica* dei domini che sta mettendo a confronto. Tutta e solo l'informazione necessaria viene fornita a SME. Ogni informazione aggiuntiva irrilevante (ad esempio, il fatto che i pianeti possono avere satelliti) non fa parte della rappresentazione nel calcolo dei predicati che viene fornita al modello. La distinzione fra relazioni, attributi e oggetti proposta dalla Gentner diviene parte costitutiva, rigidamente incorporata, della rappresentazione fornita al programma, anche se, come fanno notare Chalmers, French e Hofstadter, nella loro discussione del modello, «dal punto di vista psicologico, molti concetti appaiono oscillare tra la qualifica di oggetto e quella di attributo. [...] Perciò, quando si progetta una rappresentazione da sottoporre a SME, si deve operare un buon numero di scelte arbitrarie, ciascuna delle quali influisce in misura rilevante sulle prestazioni del programma» (Chalmers, French, Hofstadter, 1992, pp. 200-201).

Il modo in cui l'elaborazione è effettuata da SME ricorda quello di ANALOGY, anche se con due importanti differenze che rendono, paradossalmente[17], il secondo psicologicamente più plausibile del primo. Innanzitutto, il fatto che ANALOGY costruisce le proprie rappresentazioni della situazione costituisce un primo passo verso l'autonomia del programma dal programmatore, nel senso che l'elaborazione risulta in questa maniera più creativa e meno guidata da schemi di rappresentazioni preconfezionate secondo le modalità e i contenuti di conoscenza di chi fornisce i dati al programma. In secondo luogo, la possibilità di costruire le proprie rappresentazioni, negata a SME, è strettamente correlata con la natura di queste rappresentazioni. Infatti, mentre il dominio di ANALOGY è quello della geometria, il che conferisce al programma la possibilità di avere una conoscenza esaustiva degli oggetti della rappresentazione, visti come entità ideali codificate attraverso un numero preciso e definibile di caratteristiche, l'obiettivo di SME è quello di scoprire analogie fra domini del mondo reale, la cui "conoscenza" va molto al di là della semplice rappresentazione strutturata nella forma del calcolo dei predicati con la quale SME è equipaggiato. Diverso sarebbe stato il caso in cui anche questo programma fosse stato dotato di un meccanismo di costruzione di rappresentazioni a partire dai fatti conosciuti nei domini che mette in relazione, fatti che, però, hanno una complessità ben maggiore di quelli che riguardano le relazioni fra semplici figure definibili per la loro forma e posizione reciproca. A SME, in definitiva, *manca la comprensione* di quello che sta trattando in misura ben maggiore rispetto ad ANALOGY e tale differenza riguarda due aspetti, uno relativo all'architettura del programma e uno al dominio di applicazione. Ad entrambi questi problemi, d'altra parte, i modelli dell'approccio subcognitivo tentano di porre rimedio.

[17] La paradossalità è dovuta al fatto che ANALOGY, presentato nel 1968, fu progettato e sviluppato negli anni in cui l'indirizzo di ricerca noto come Scienze Cognitive non era ancora nato e l'idea di costruire modelli plausibili dal punto di vista psicologico era solo una delle possibilità implicite, in molti casi considerata un proposito collaterale, dell'IA. Il fatto che si siano scelti come dominio applicativo del modello test utilizzati negli studi psicologici sulle capacità intellettive non deve trarre in inganno. La ricerca di una prestazione psicologicamente plausibile non è ancora, negli anni sessanta, unita al tentativo di costruire architetture computazionali esplicative dal punto di vista cognitivo, tentativo che comincia ad essere perseguito in modo sistematico nel corso degli anni settanta con la nascita del paradigma HIP (*Human Information Processing*).

In seguito sono state presentate alcune varianti di SME. Una delle più interessanti è quella che prevede l'estensione di questo programma con un modello di recupero basato sulla somiglianza fra situazione presente e episodi conservati in memoria. Tale modello, denominato MAC/FAC (Forbus, Gentner, Law, 1995), si avvale di una memoria a lungo termine in cui sono presenti, e già formalizzati nella logica dei predicati, eventi passati. Il recupero viene effettuato in due fasi. La prima è la ricerca di episodi in memoria, attraverso somiglianze superficiali tra situazioni, per mezzo della corrispondenza istituita fra predicati condivisi. Una volta determinati gli episodi più vicini alla situazione obiettivo, il migliore viene selezionato attraverso l'utilizzo estensivo delle rappresentazioni delle situazioni. Si colmano, così, due lacune di SME, attraverso la rinnovata attenzione alle somiglianze superficiali unitamente a quelle strutturali, in ragione dell'assunzione che entrambe rivestono la stessa importanza nel processo analogico, e grazie alla presenza di una memoria episodica che fa sì che il modello possa scegliere fra diverse situazioni sorgente e non si limiti a costruire diverse possibilità di mappatura, tra cui scegliere, fra le rappresentazioni di due domini soltanto. Il modello MAC/FAC può essere fatto rientrare nell'impostazione del *case-based reasoning* proposta da Schank, come soluzione, non scevra da problemi, alla più generale questione della comprensione di un dominio.

2.3.2 Modelli connessionisti

I nodi irrisolti dell'approccio simbolico al ragionamento analogico hanno spinto negli ultimi anni a tentare nuove impostazioni di ricerca. Nel caso dell'analogia, ancor più che in altri tentativi di simulazione dei processi cognitivi, è possibile apprezzare il modo in cui a un medesimo problema il connessionismo dà una differente soluzione che deriva direttamente dai suoi fondamenti teorici costitutivi. Se, infatti, per quanto riguarda l'approccio simbolico era relativamente semplice mettere in corrispondenza, attraverso l'operazione di *matching*, due situazioni sulla base delle loro caratteristiche *identiche*, una rete connessionista si presta molto meglio a cogliere le somiglianze fra entità non basate su relazioni di uguaglianza. Semplificando, si può affermare che ciò derivi dal modo in cui una rete codifica in maniera distribuita l'informazione in input e riproduce in output lo stesso, o quasi lo stesso, schema di attivazione a fronte di un input simile a quelli dello stesso tipo codificati in precedenza durante il processo di addestramento.

È noto che nei compiti di riconoscimento i modelli connessionisti abbiano dato risultati migliori rispetto a quelli simbolici sia dal punto di vista della prestazione che da quello del rapporto fra risorse computazionali utilizzate e risultati conseguiti. Tuttavia, come si è visto, il ragionamento analogico eccede i confini dei processi di categorizzazione e riconoscimento e si sono dovuti escogitare metodi per modellizzare anche i processi di mappatura fra situazioni e domini differenti. Uno dei modelli più conosciuti è Analogical Constraint Mapping Engine (ACME) di Holyoak e Thagard (1989) basato sul principio secondo cui un'analogia è il risultato della soddisfazione

complessiva di un insieme di vincoli, raggiungibile attraverso l'elaborazione compiuta da una rete connessionista in cui a ogni nodo corrisponde una coppia di elementi rispettivamente della rappresentazione della situazione nota e di quella obiettivo. Le connessioni della rete, legami pesati, costituiscono i vincoli strutturali del sistema e permettono la diffusione e il reciproco supporto dell'attivazione fra nodi che esprimono ipotesi consistenti con la mappatura, e l'inibizione dei nodi non rilevanti per il processo di mappatura. Riprendendo l'esempio di Hitler e Caporetto, si può pensare che vengano creati nodi-coppie come "Stalingrado → guerra di invasione" o "Stalingrado → sconfitta", i quali ricevono attivazione, ma anche nodi come "Stalingrado → battaglia di trincea", che, al contrario, vengono inibiti. Lo stato di equilibrio viene raggiunto attraverso un algoritmo di rilassamento della rete. In questo modo si arriva alla mappatura migliore, ovvero all'insieme di tutti i nodi attivati che esprimono l'ipotesi più adeguata di accoppiamento.

È possibile attuare una procedura di recupero attraverso la diffusione di attivazione nella rete[18]. Il risultato viene comunicato sottoforma di legami fra predicati e argomenti, così come del resto era stato fornito l'input. Anche ACME funziona, dunque, attraverso rappresentazioni rigide e costruendo, seppure in parallelo, tutti gli accostamenti sintattici possibili al fine di scoprire quelli che costituiscono il migliore accoppiamento complessivo. La novità di questo modello rispetto ai precedenti è la presenza di due nodi che esprimono rispettivamente la somiglianza semantica e la rilevanza pragmatica dei nodi rappresentanti le coppie di predicati. Tuttavia, come fanno notare Mitchell e Hofstadter (1994), la valutazione della somiglianza semantica è fatta dal programmatore, così come quella dell'importanza a livello pragmatico dei nodi coppie per una mappatura efficace. Di conseguenza, la comprensione semantica dei domini da parte del programma è anche nel caso di ACME limitata alla conoscenza delle relazioni formalizzate preconfezionate e nelle valutazioni del programmatore; e, per tale motivo, si può considerare in larga misura assente.

2.3.3 Modelli ibridi

In questo tipo di approccio si cercano di sfruttare i punti di forza degli altri due nel progettare e implementare modelli del ragionamento analogico. Perciò, mentre i processi di alto livello, come quello di mappatura fra due situazioni, vengono realizzati con metodi simbolici, cioè attraverso il confronto e la messa in corrispondenza biunivoca di relazioni fra elementi, la somiglianza fra gli elementi stessi, che non necessariamente collassa in uguaglianza, è ottenuta attraverso l'impiego di metodi connessionisti che rappresentano l'informazione in forma distribuita, non rigida e sensibile al contesto. Ancora, come fanno notare Kokinov e French (2003, p. 116), «mentre la proiezione per mappe è guidata dalla somiglianza di strutture, il recupero è guidato dalla somiglianza semantica», cosicché i metodi connessionisti e simbolici trovano ognuno una propria collocazione. In termini

[18] Il recupero dell'informazione memorizzata attraverso l'"immissione" di attivazione in una rete già addestrata è una caratteristica peculiare dei modelli connessionisti in genere.

generali, si può dire che i modelli ibridi implementano, o cercano di farlo, l'analogia *lato sensu*, ovvero sia come processo di riconoscimento e categorizzazione, che come proiezione di una struttura di relazioni "profonda" da una situazione sorgente a una situazione obiettivo.

Inoltre, una delle caratteristiche di questi modelli è l'avvalersi di un tipo di elaborazione parallela, eseguita da una serie di microprocedure che agiscono sulla base di parametri probabilistici. È il caso, ad esempio, di Associative Memory-Based Reasoning (AMBR) di Kokinov (1994), basato su un'architettura di questo tipo denominata DUAL. Ogni nodo della rete rappresenta una delle microprocedure che viene "chiamata" nel momento in cui la quantità di attivazione, dovuta all'attività di propagazione nella rete, supera un certo valore di soglia. L'insieme dei nodi più attivi costituisce la rappresentazione dei concetti e degli episodi coinvolti nella analogia, i quali vengono recuperati dalla memoria episodica distribuita e incorporata nella rete. Tuttavia, l'impiego di tale metodo per la costruzione di analogie fra situazioni del mondo reale lo rende simile agli altri quanto al contenuto semantico rappresentato, il quale viene immesso globalmente dal programmatore e non è lasciato in definitiva all'elaborazione del programma. Ritroviamo anche in questo caso il solito problema delle rappresentazioni preconfezionate.

Altri modelli godono delle stesse caratteristiche di architettura, pur con alcune particolarità. L'evoluzione del modello di Kokinov, AMBR-2 (Kokinov, Petrov, 2001), introduce una memoria di lavoro che ricostruisce la situazione data in input attraverso l'informazione contenuta nella memoria a lungo termine e la confronta con la rappresentazione iniziale. I modelli STAR sfruttano rappresentazioni espresse sottoforma di prodotto tensoriale (*à la* Smolensky) a tre o quattro dimensioni. Nel modello LISA viene introdotto un ulteriore parametro, quello temporale, secondo il quale vengono misurate le oscillazioni negli schemi di attivazione. Una sincronia nell'oscillazione dell'attività di schemi diversi sta a indicare un loro allineamento analogico[19].

Come si vede, il problema di modellizzare il ragionamento analogico ha interessato tutti i principali approcci in IA e nelle scienze cognitive. Esso coinvolge molti aspetti cruciali della ricerca: dalla categorizzazione, alla natura dei concetti, alla rappresentazione della conoscenza, configurandosi quasi come una specie di banco di prova, di cartina al tornasole, da una parte per le metodologie e i problemi e dall'altra per gli assunti teorici tipici di ogni approccio. Anche i modelli subcognitivi possono essere fatti rientrare – si veda la classificazione dei modelli del ragionamento analogico di Kokinov e French (2003) – fra quelli ibridi, anche se più per il fatto di condividere un'impostazione che tenta di unificare processi *bottom up* e processi *top down*, che non per il fatto di incorporare moduli tipicamente connessionisti nella loro architettura. D'altra parte, non tutti gli autori di questi modelli riconoscono la loro natura ibrida e, nel proporre un tipo alternativo di architettura, ne individuano la sua caratteristica principale nel fatto di essere *emergente, ma non*

[19] Per una panoramica su tutti questi moelli si rimanda a Kokinov e French (2003). Sul modello LISA, che introduce il fattore temporale, principalmente come vincolo psicologico, si rimanda a Hummel e Holyoak (1997).

connessionista. Sul come e sul perché ciò avvenga si discuterà a lungo nel prossimo capitolo, in cui verranno delineate le differenze e le peculiarità dei vari modelli sviluppati in questo ambito di ricerca.

Qui di seguito ci accingiamo a trattare la seconda loro principale caratteristica: i microdomini di applicazione.

2.4 La questione dei microdomini

Uno dei tratti che forse rende più impopolari e meno conosciuti i modelli realizzati dal *Fluid Analogies Research Group* (FARG) riguarda la propensione alla scelta per i programmi di domini ristretti di applicazione. Tali domini vengono, appunto, chiamati "microdomini" per distinguerli dai più conosciuti micro-mondi che tanta parte hanno avuto nella storia dell'IA, soprattutto sul versante critico dell'IA stessa. Tuttavia, un punto di contatto fra le due nozioni esiste e riguarda la scelta del livello di complessità che un programma è in grado di affrontare. Utilizzare il mondo reale come dominio di applicazione vuol dire condannare all'immobilità taluni modelli, vista la complessità delle informazioni che dovrebbero entrare in gioco nel processo computazionale, o che anche soltanto dovrebbero essere in qualche maniera immagazzinate nella memoria. Un'ampia ed estesa base di conoscenza è certo sempre inferiore alla descrizione completa del mondo reale, sia per via della sua rigidità e staticità se si parla di rappresentazioni in forma simbolica, rigidità e staticità dovute alla giustapposizione di lunghissimi elenchi di enunciati che non possono esprimere tutte le relazioni dinamiche del mondo reale; sia, se si parla specialmente delle rappresentazioni distribuite dei modelli connessionisti, per la difficoltà di recupero dell'informazione immessa nella rete, difficoltà che cresce in maniera proporzionale alla quantità dell'informazione, visto che la rete viene ricalibrata per ogni elemento informazionale con un rischio crescente di interferenza catastrofica, cioè la perdita dell'informazione già appresa e immagazzinata in forma distribuita nella rete[20].

D'altra parte, sono note le critiche che vennero portate all'impiego di micro-mondi fin dall'apparizione dei primi sistemi simulativi di comprensione del linguaggio naturale, come SHRDLU di Winograd (Winograd, 1972), o dei primi tentativi di programmi impiegati in compiti di costruzione categoriale, come il modello, sviluppato da Winston, di apprendimento per mezzo di esempi (Winston, 1975). Per Dreyfus, ad esempio, i micro-mondi non colgono la complessità del mondo reale perché sono modelli astratti che nulla hanno a che vedere con il mondo. Infatti, pur delimitandone una parte, non ne diminuiscono la complessità, che rimane la stessa di quel mondo preso nella sua interezza che essi presuppongono. Perciò, un micro-mondo, come quello dei blocchi

[20] L'interferenza catastrofica è un problema tanto più grande quando maggiore è la distribuzione della rappresentazione della conoscenza nella rete. Una possibile via di uscita è costituita dal localismo rappresentazionale, che, però, indebolisce uno dei punti di forza del connessionismo, la robustezza e la flessibilità dell'informazione codificata in maniera distribuita.

geometrici solidi in cui agisce SHRDLU è un dominio *preconfezionato* e «un insieme di fatti interconnessi può costituire un *universo*, un dominio, un gruppo, ecc., ma non costituisce un *mondo*, perché un mondo è una quantità organizzata di oggetti, scopi, abilità e pratiche secondo cui le attività umane hanno significato [...]. Se i micro-mondi *fossero* sotto-mondi, non ci sarebbe bisogno di elaborarli e combinarli per avvicinarli al mondo quotidiano, poiché quest'ultimo sarebbe già incluso» (Dreyfus, 1981, p. 184-185). Un discorso analogo è valido per i modelli di estrazione categoriale, che ricadono all'interno dell'ambito del *machine learning* e che si avvalgono di una serie di primitive selezionate dal programmatore per la costruzione di descrizioni formalizzate di determinate categorie[21].

La discussione sui micro-mondi è, dunque, strettamente intrecciata a quella sulla rappresentazione delle conoscenze e al tentativo di superare la rigidità dei sistemi di conoscenza immessa nei programma di IA degli anni Settanta, cioè nel periodo in cui cominciarono ad apparire i primi programmi in grado di affrontare compiti di una qualche rilevanza con *prestazioni* per certi aspetti uguali se non superiori a quelle umane dovute proprio alla gran quantità di conoscenza immagazzinata. Tali programmi, che furono denominati sistemi esperti e si distinsero, ad esempio, nel campo della diagnostica medica, spinsero i filosofi e i teorici dell'IA a interrogarsi sulla vera natura della comprensione di tale conoscenza, fino ad arrivare a conclusioni che, pur affermando il valore e l'utilità di tali applicazioni ai fini pratici, vedevano in esse una quasi totale deviazione dalle peculiarità della conoscenza *human-like*. In altri termini, i sistemi esperti vennero considerati molto poco plausibili dal punto di vista psicologico. D'altra parte, la ragione per cui vennero ideati e costruiti non aveva come obiettivo primario un intento esplicativo cognitivo. Il loro muoversi in un dominio specifico in cui la conoscenza è totalmente strutturata secondo i metodi del calcolo dei predicati non ne faceva dei veri *conoscitori* di quel dominio, ma solo, per così dire, dei supporti attivi per l'utente, ad esempio per il personale medico, che necessita di conclusioni esatte al termine di un processo deduttivo condotto su una base molto ampia, poco dominabile da una mente umana, di premesse, come ad esempio l'insieme dei sintomi e delle malattie corrispondenti in uno specifico settore della medicina.

Tuttavia, i micro-mondi e i microdomini differiscono per almeno un aspetto fondamentale. Tale differenza si gioca sul tipo di capacità che il programma dovrebbe simulare. Infatti, SHRLDU aveva come obiettivo quello di modellare la capacità di comprensione del linguaggio naturale e, quindi, del mondo che tale linguaggio esprimeva. Tuttavia, poiché si trattava di un universo costruito *ad hoc*, anche la comprensione che ne derivava e che veniva espressa attraverso un dialogo in linguaggio naturale, era soltanto parziale e fittizia, o, quantomeno ingannevole, perché realizzata

[21] Nel modello proposto da Winston si fa l'esempio della costruzione della categoria di arco attraverso l'enucleazione delle sue caratteristiche principali, a partire da una serie di *item* sottoposti all'elaborazione del programma, individuate per mezzo di un insieme di proprietà e relazioni basilari di cui esso è dotato in partenza (ad esempio, proprietà: oggetto (pilone), oggetto (trave), azione (passare sotto), azione (passare attraverso), ecc.; relazioni: sopra(x,y), a destra di(x,y), è sostenuto da(x,y), ecc.)

dalle componenti procedurali di cui il programma era composto[22] e sicuramente inferiore a quella che l'utente era portato ad attribuirgli. Il fatto di utilizzare un universo *ad hoc* era l'espediente attraverso cui si riduceva la complessità del mondo reale ad una trattabilità che rendeva, però, parimenti priva di profondità anche la nozione di comprensione attribuita al programma, se riferita al linguaggio naturale nei suoi aspetti semantici più generali che lo rendono tutt'uno con la complessità del mondo reale che esprime.

Detto altrimenti, la presenza di un modello idealizzato e ristretto del mondo non permette di parlare di una comprensione del significato simile a quella umana, anche nel caso di un riferimento al medesimo universo, da parte del programma, se la dimostrazione di questa comprensione viene cercata e testata nell'interfaccia in linguaggio naturale. Perciò, mentre si può dire che molta parte del *background* del mondo viene eliminato prendendo in considerazione un certo particolare universo, esso non può essere ugualmente tolto dal linguaggio naturale che *funziona* proprio sulla base, e per la presenza, di questo *background*[23]. In definitiva, da una parte il mondo dei blocchi è quella parte di mondo, ridotto a universo specifico, che il programma conosce bene e che collassa sulla nozione di microdominio; dall'altra il linguaggio naturale esprime un mondo che non può essere parzializzato, a meno di ricorrere a formalizzazioni che ne mutino la natura e le possibilità, e di cui, pertanto, non ha senso, o è fuorviante, parlare di conoscenza parziale, micro-contestualizzata.

I microdomini hanno, dunque, la funzione specifica di non ingannare il programmatore o l'utente del programma in merito alle reali conoscenze del programma, e, perciò, di non instillare l'idea di una comprensione *human-like* da parte del programma. I modelli che li utilizzano, quando si avvalgono di un'interfaccia in linguaggio naturale, non lo fanno all'interno di un dialogo con l'utente, ma solo per rendere più facilmente comprensibile all'utente la regola analogica che soggiace alle diverse entità considerate. Il microdominio serve, perciò, a esplicitare l'*ignoranza* del programma, piuttosto che la sua *conoscenza*, la quale, invece, deve essere facilmente individuabile nell'architettura che si può, eventualmente a ragione, affermare simulativa del processo attivo di comprensione. È sulla ambiguità di questo termine che sono state costruite le critiche all'IA a causa dell'elusiva aura di impalpabilità che sprigiona, come è palese nelle affermazioni di Dreyfus e ancora di più in quelle di Searle.

Tutto ciò viene meno nel momento in cui una teoria delle comprensione è data in senso positivo, ovvero per quanto riguarda i meccanismi che possono generarla, e non solo in senso negativo, cioè per mezzo dell'affermazione di ciò che la comprensione *non è*. Infatti, una volta ipotizzata una teoria dei processi di comprensione, se ne può tentare una verifica attraverso la costruzione di un

[22] Nel caso specifico del programma di Winograd si tratta di un analizzatore sintattico delle frasi immesse dall'utente, una base di conoscenze e un sistema deduttivo per trarre inferenze a partire dalle conoscenze implementate in forma predicativa.

[23] Un discorso analogo può essere fatto per i modelli di estrazione categoriale, se il processo di *analisi* viene effettuato sì dal programma, ma attraverso un insieme rigido e *non modificabile dal programma* di relazioni e proprietà primitive. Per avere un buon modello, un modello *significativo*, di *machine learning* occorre che l'apprendimento non riguardi soltanto le situazioni sussunte sotto certe categorie, ma i modi stessi di categorizzare, almeno come traguardo finale da conseguire.

modello, il cui dominio di applicazione deve avere una larghezza ampia *almeno* fino al punto di rendere possibile il funzionamento del modello, ma *non necessariamente* tanto grande da includere una più illusoria che reale comprensione di tutte le cose. Lo studio dei fenomeni attentivi in psicologia sperimentale ha mostrato, peraltro, che le capacità umane attingono a risorse limitate e, in genere, presuppongono una qualche forma (ancora discussa) di filtraggio percettivo. Ciò fa sì che la memoria a breve termine abbia a che fare con un numero esiguo di elementi (individuato da Miller nel famoso "numero magico" 7 ± 2), e si applichi di preferenza, anche se non necessariamente, a un dominio specifico, rimandando alla definizione della natura del rapporto con la memoria a lungo termine la questione del recupero del dominio specifico adatto ai compiti oggetto dell'attenzione cosciente. Tali limitazioni dovrebbero essere incorporate anche in un modello che voglia spiegare i meccanismi alla base del ragionamento analogico. I microdomini acquistano di conseguenza il ruolo di idealizzazioni tali da permettere la verifica in più ambiti, tanti quanti sono i microdomini cui i modelli vengono applicati, dei risultati sperimentali della psicologia.

In questa ottica va letta l'affermazione di French, secondo il quale, per quanto riguarda i microdomini, non è fuorviante un'analogia con la sperimentazione in fisica:

> Si consideri il modo in cui la fisica è progredita. Per studiare le proprietà e il comportamento della materia in movimento, Newton fece grandi passi avanti trattando i corpi nello spazio come punti e ignorando la nozione di attrito. I progressi nella fisica, sia che abbiano riguardato lo studio dei gas, dell'elettricità, del calore o delle particelle subatomiche, sono sempre stati dipendenti dall'uso di modelli idealizzati. Le idealizzazioni sono utilizzate in modo tale che non si debbano tenere in considerazione, almeno inizialmente, le numerose influenze che potrebbero mascherare l'investigazione delle proprietà essenziali. Una volta che tali proprietà siano state descritte, i vincoli dell'idealizzazione nel sistema possono gradualmente essere allentati così da permettere lo studio del problema in un ambiente più generale, allo scopo di perfezionare il modello. (French, 1995, p. 23)

Il processo di idealizzazione è, dunque, necessario alla messa in risalto del fenomeno in oggetto e la scelta di un modello ideale della realtà *non è* strettamente dipendente dal fatto che esso venga utilizzato come dominio di un modello cognitivo, a meno di non voler perdere la generalizzabilità di quest'ultimo e della teoria che esso implementa come spiegazione di un'intera classe di fenomeni (cognitivi). In effetti, come fa notare la Mitchell, non è per il fatto che ci si avvicini a un dominio artificiale e non al mondo reale, che gli esseri umani smettano di utilizzare quei

> [...] meccanismi percettivi che si sono evoluti nel continuo commercio con le situazioni reali nel mondo reale. Questi meccanismi non si accendono o si spengono semplicemente perché il dominio è

> apparentemente artificiale e indipendente dal contesto e perché la nostra sopravvivenza non dipende dalle nostre azioni [in quale dominio]. (Mitchell, 1993, p. 26)

Ciò costituisce uno degli argomenti principali a sostegno dell'impresa metodologica della ricerca in psicologia e può anche essere considerato una sorta di scelta di campo diversa rispetto all'assunzione di una certa parte dell'IA, la quale vede nella parcellizzazione e nella semplificazione del dominio l'impossibilità di considerare tali modelli come genuine simulazioni dei processi di *problem solving* rivolti allo scopo tipici del mondo reale. Tuttavia, anche in questo caso sembra essere coinvolta una certa ambiguità, che si era vista in precedenza nel caso della nozione di comprensione e che adesso può essere meglio specificata come sovrapposizione di due elementi distinti: da una parte il *fenomeno* simulato, dall'altra i *meccanismi* ipotizzati per la spiegazione del fenomeno. Come molti argomenti critici dell'IA sono attribuibili alla fusione indebita fra *atto* di comprensione e *processo* di comprensione, così ugualmente le critiche ai microdomini possono essere dettate da una confusione fra dimensione (ristretta rispetto al mondo reale) del dominio utilizzato e dimensione (allargata rispetto al dominio) dei fenomeni cognitivi indagati. Perciò, come si vedrà in seguito, sia che si tratti di domini chiusi, cioè con un numero finito di elementi (ad esempio, l'alfabeto), sia che si tratti di domini aperti, in cui il numero degli elementi è potenzialmente illimitato (ad esempio, l'insieme dei numeri naturali), occorre sempre tener presente *ciò che è in gioco nella simulazione*, ovvero il fatto che tali microdomini intendono

> [...] essere strumenti per esplorare gli aspetti generali [– la fluidità concettuale che permette la percezione di alto livello –] della cognizione piuttosto che quelli specifici dell'ambiente di lettere e stringhe, o quelli di domini ristretti a strutture lineari con distanze note tra gli elementi. (Mitchell, Hofstadter, 1994, p. 229)

A questo punto una possibile obiezione potrebbe essere relativa alla specificità del dominio rispetto al modello cognitivo, *nella misura in cui* il modello appare essere progettato con caratteristiche *ad hoc* per il dominio opzionato. Questo problema è un problema epistemologico e si riallaccia alle assunzioni che vengono fatte nella teoria e *in base alle quali* il modello viene progettato. Come si vedrà in seguito, l'architettura dei modelli varia a seconda dei (micro)domini di applicazione, ma esiste, o dovrebbe esistere, un nucleo architettonico comune a tutti, che si suppone implementi il meccanismo essenziale della teoria (la percezione di alto livello) e che i modelli applicano nei vari domini attraverso opportune variazioni costruttive "superficiali". Esse permettono al modello di operare in quel determinato dominio e consistono, ad esempio e in linea del tutto generale, in variazioni nei concetti della memoria semantica, nelle possibilità della memoria di lavoro, nei micro-algoritmi applicativi della memoria procedurale effettiva, e così via. Una valutazione dell'efficacia predittiva dei modelli e, quindi, una convalida del nucleo della loro

architettura sarà tentata più avanti nel corso di questa trattazione. Ora, conviene passare all'ultimo dei loro tratti distintivi, che consiste appunto nello schema di base della loro architettura.

2.5 L'architettura cognitiva dei modelli

In termini generali, lo scopo dichiarato dei modelli subcognitivi che operano nei microdomini è quello di capire non *che cosa* un programma comprende, ma *come* un programma comprende, ovvero implementare un'architettura che è alla base del *processo* di comprensione, attraverso la simulazione del meccanismo della fluidità concettuale, sotteso alla percezione di alto livello, che costituisce, a sua volta, il nucleo dei processi analogici *lato sensu*. Strettamente parlando, l'idea di fondo è che per capire che cosa un modello cognitivo comprende occorre capire in che modo può essere messo in grado di comprendere. Considerare le cose da questo punto di vista indebolisce la questione del microdominio fino a riportarla alla valutazione dell'adeguatezza del modo in cui vengono condotti gli esperimenti in psicologia, la cui metodologia della ricerca prevede prevalentemente una circoscrizione della prestazione analizzata a domini ristretti e idealizzati al fine di verificare o falsificare gli assunti principali di una teoria o di un insieme di teorie.

La domanda sulla validità di tale approccio si sposta, di conseguenza, sugli assunti della teoria e sulle modalità della sua implementazione algoritmica. Vale a dire, che tipo di architettura impiegano questi modelli ai fini della simulazione della capacità di comprensione? Il che equivale a chiedersi: quali meccanismi permettono la percezione di alto livello? Per rispondere a queste domande è stata formulata, all'interno dell'approccio subcognitivo, una teoria dei processi di ragionamento che coinvolge processi di memoria, di focalizzazione attentiva, di elaborazione dell'informazione e di organizzazione e creazione di nuova conoscenza concettuale. Tale teoria è stata recentemente enunciata, in termini generali, da Hofstadter e potrebbe essere esplicitamente definita dalle seguenti parole:

Teoria dell'«*Anello Centrale della Cognizione: un nodo della memoria a lungo termine è penetrato, trasferito nella memoria a breve termine e là spacchettato a un qualche grado, il che fa sì che nuove strutture vengano percepite, e l'atto percettivo di alto livello, che così si produce, attivi ancora ulteriori nodi, che sono a loro volta penetrati, trasferiti, spacchettati, ecc., ecc.*» (Hofstadter, 2001, p. 517)

Mettere alla prova questa teoria è possibile attraverso la costruzione di modelli che ne rispecchino le caratteristiche. Nel fare questo, appaiono immediatamente evidenti due cose: in primo luogo, che tale teoria implica un sistema composto di parti diverse in interazione; in secondo luogo, che l'interazione deve essere continua e virtualmente infinita, nel senso che è l'*anularità* del

modello, ovvero la possibilità di una applicazione ricorsiva di elementi del sistema all'informazione trattata da altri elementi del sistema, a rendere tale il processo di comprensione, cioè di percezione di alto livello. Un punto di arresto del sistema, che in un programma di IA corrisponde generalmente al momento della comunicazione della soluzione, corrisponde alla cessazione del *loop* cognitivo e quindi anche alla fine del processo di comprensione. Con queste affermazioni non si vuole dire che il programma "muore" in un qualche senso del termine. Da questo punto di vista, ogni programma condivide un simile destino. D'altra parte, una delle differenze fondamentali fra l'uomo e la macchina, che troppo spesso viene offuscata e messa in ombra da quella relativa al materiale di cui sono diversamente formati, è che l'esecuzione di un programma è a termine – e, se non lo fosse, ciò significherebbe che il programma non funziona a causa di un errore algoritmico (e non semplicemente di codice) – pure se l'*hardware* su cui gira continua a rimanere acceso, mentre questa fase di *stand by* non è concessa ad un essere umano. Se lo fosse, se cioè si riuscisse a dare una definizione di *stand by* valida anche per l'essere umano, ciò probabilmente inciderebbe fortemente sulla visione di radicale differenza fra le conoscenze possedute da un sistema artificiale e quelle possedute da un sistema umano, ridimensionandola. In ogni caso, la fine del processo di comprensione va piuttosto considerata nella prospettiva secondo il processo di ricerca di una soluzione è più importante della soluzione stessa ai fini della spiegazione di che cosa il modello comprende, anche in considerazione del fatto che è dall'interazione fra le sue parti che dipende buona parte della rappresentazione della conoscenza posseduta dal programma a un certo istante di tempo t_n. Tuttavia, è necessario che il programma si arresti, ovvero che *arrivi a considerare il processo di comprensione sufficiente per la consegna di una soluzione*.

Il ricorso a diversi tipi di memoria della Teoria dell'Anello Centrale della Cognizione (*Theory of Central Cognitive Loop*, d'ora in avanti TCCL) riprende la distinzione classica in psicologia cognitiva fra Memoria a Lungo Termine (MLT) e Memoria a Breve Termine (MBT), della quale sono stati presentati più modelli. Il primo risale a Atkinson e Shiffrin (1968) e prevede tre moduli: un registro sensoriale, che trattiene l'informazione per 3-4 secondi; la MBT, con capacità limitata relativamente agli elementi che può contenere (la quantità individuata da Miller, cioè 7 ± 2) e alla durata del trattenimento (dai 20 ai 30 secondi circa); e la MLT, i cui limiti di capacità non sono individuati con precisione e dunque sono ritenuti virtualmente illimitati[24]. Inoltre, tale modello prevede l'interazione reciproca fra MBT e MLT.

Baddeley (1986) lo ha perfezionato introducendo una suddivisione in tre sottomoduli all'interno della MBT. Essa sarebbe composta, infatti, da un sistema esecutivo centrale collegato a due

[24] Questo fatto non è necessariamente in contraddizione con la limitatezza del supporto su cui la memoria viene realizzata, il cervello, se si ipotizza che la MLT possa consistere in un meccanismo ricorsivo di attivazione neurale. Tuttavia, non c'è ancora un'evidenza sperimentale in proposito o un pieno accordo nell'interpretazione dei risultati sperimentali a livello di singole cellule o di insiemi di cellule del cervello. Rimane il fatto che ispirarsi a meccanismi neurali di questo tipo costituisce una delle vie attraverso cui la computazione cerca di superare i limiti imposti dalla Turing-computabilità. Su questo tema, e sui suoi aspetti matematici, si rimanda all'accurato studio della Siegelmann (1999).

sottosistemi tra loro indipendenti: il *loop* articolatorio, adibito all'elaborazione e al mantenimento dell'informazione linguistica, e il taccuino visuo-spaziale, implicato nell'elaborazione e nel mantenimento dell'informazione rilevante dal punto di vista spaziale[25].

Anche per la MLT sono state proposte sottoparti specifiche, ad esempio da Tulving (1972), che nel suo modello ipotizza una memoria procedurale, una memoria semantica, e una memoria episodica. La prima contiene schemi o sequenze di azioni *goal-oriented*; la seconda le relazioni fra i concetti e i concetti stessi; nella terza sono allocate le specifiche esperienze passate del soggetto. Anche questi tre sottotipi di MLT sono organizzati in maniera gerarchica. La memoria procedurale è considerata la più basilare per l'astrattezza e la generalizzazione delle informazioni che contiene. Le altre due sono sottomemorie collegate alla prima con un diverso grado di specializzazione. La memoria semantica, infatti, gode di un grado di generalizzazione maggiore rispetto alla memoria episodica ed è, dunque, più vicina a quella procedurale.

Molti modelli della memoria proposti nel corso degli anni sono stati inclusi per alcune o altre caratteristiche nei programmi di IA prodotti dall'approccio subcognitivo. In termini generali, si può affermare che nella TCCL assume un notevole peso la questione della memoria, suddivisa fra i due sottotipi immediatamente meno specifici, la MLT e la MBT. Questo è un tentativo di incorporare nella teoria due intuizioni: che non c'è rappresentazione senza una qualche pur minima forma *articolata* di memoria e che l'*interazione* fra due o più memorie è la chiave di volta per la spiegazione dei fenomeni di creazione di analogie descritti in precedenza, compresi nell'ampio spettro che va dal riconoscimento, alla categorizzazione, alla rievocazione di eventi e situazioni, all'analogia vera e propria, cioè *stricto sensu*. Quale è, dunque, la natura di tale interazione e come può essere descritta e inserita nel modello?

Nel 1982 Roger Schank ha proposto un perfezionamento della sua teoria degli *script* formulata negli anni Settanta. Per superare il problema relativo alla rigidità e alla staticità della conoscenza implementata nei programmi di comprensione di racconti espressi in linguaggio naturale, quale era per esempio SAM[26], egli formulò un modello di *memoria dinamica* definito come «sistema flessibile finito-aperto» (Schank, 1982, p. 9). Tale modello doveva costituire un punto di contatto, migliore di quelli proposti in precedenza, fra i meccanismi alla base dell'utilizzo della memoria umana e la loro implementazione in IA, superando alcuni dei problemi classici relativi alla rappresentazione della conoscenza in un programma. Schank ipotizzò che vi fosse una forma di memoria intermedia fra la MBT e la MLT e che essa permettesse il passaggio selettivo di

[25] Tale specificazione della MBT è stata proposta in seguito alla contrapposizione, a cavallo fra gli anni settanta e gli anni ottanta, fra immaginisti e proposizionalisti in merito al modo in cui l'informazione è immagazzinata nel pensiero e aspira a essere una plausibile spiegazione integrata di entrambi i fenomeni. Per i termini e gli sviluppi storici di tale disputa si rimanda a Luccio (1998).

[26] SAM è descritto approfonditamente in Cullingford (1978). Esso costituisce un'applicazione della nota Teoria della Dipendenza Concettuale di Schank (1972) che ha portato alla formulazione della Teoria dello *Script*, sulla base della quale, e contro la quale, Searle ha costruito l'argomento della stanza cinese.

informazioni dalla prima alla seconda. Infatti, solo in questo modo si spiegherebbe perché la MLT non trattiene tutta l'informazione disponibile, ma solo una parte di essa, cioè quella più rilevante per l'individuo. La natura della comprensione è legata strettamente a questo tipo di processo e, di conseguenza, essa viene fatta collassare sull'operazione di rievocazione di ricordi:

> comprendere un input significa trovare nella passata esperienza l'approssimazione più vicina all'input ed alla codificazione relativa nei termini del precedente ricordo, con un indice che indica la differenza tra il nuovo input e il vecchio ricordo. Comprendere, dunque, implica usare gli insuccessi delle aspettative guidati da ricordi prototipali o da ricordi specifici indicizzati da ricordi prototipali. *Comprendere è rievocare, e rievocare è trovare la corretta struttura di memoria per elaborare un input*. Il nostro problema più importante, quindi, nel formulare una teoria del comprendere, è di scoprire come sono le strutture di memoria ad alto livello che sono richieste. [...] Il punto chiave del comprendere è proprio questa creazione continua di nuove strutture ad alto livello, in cui vengono registrate le similarità essenziali fra esperienze differenti. (Schank, 1982, pp. 63-64 [enfasi mia])

In questa definizione entrano in gioco molti elementi. Innanzitutto, il tipo di memoria che viene chiamato in causa è di natura episodica, come era già stato nella Teoria dello *Script*. In secondo luogo, appare evidente che un fattore fondamentale consiste nell'apprendimento di nuovi episodi a partire da quelli vecchi, di norma trasformati in situazioni prototipiche di sequenze di eventi. In terzo luogo, la nozione di similarità anche in questo caso gioca un ruolo essenziale nel rendere possibile la *dinamicità* della memoria. Infine, è necessaria una struttura computazionale che implementi il livello intermedio di filtraggio fra la MBT e la MLT, la quale è individuata da Schank nel MOP (*Memory Organization Packet*). Il modello che ne deriva è quello rappresentato nella figura 2.3.

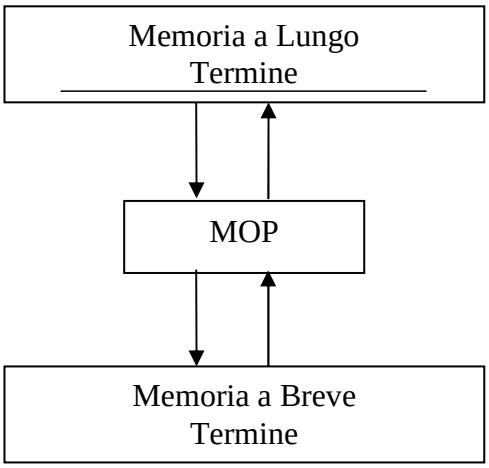

Fig. 2.3

Il MOP è allo stesso tempo una struttura di memoria, che consente di immagazzinare nuovi dati, e una struttura di elaborazione, che collega gli eventi conservati nella memoria episodica alla situazione oggetto dell'elaborazione del programma per via di somiglianza e creando aspettative di input futuri o permettendo l'inferenza di eventi impliciti. In sostanza, essi si differenziano dagli *script* e dagli scenari per la loro maggiore astrattezza. I MOP, infatti, sono ordinatori di scenari adatti a rappresentare molti sfondi diversi[27]. In tal modo, diversi MOP si adattano a diversi scenari di azione (ad esempio, il MOP-CONTRATTO e il MOP-RISTORANTE condividono lo scenario del PAGARE) così come un solo MOP si può adattare a più scenari specifici (ad esempio, il MOP-CONTRATTO si adatta agli scenari di PAGARE e di FIRMARE). A un livello di astrazione ancora maggiore rispetto ai MOP sono i TOP (*Thematic Organization Points*), strutture indipendenti dal contesto che permettono l'organizzazione coerente e la generalizzazione degli episodi memorizzati attraverso raggruppamenti binari di categorie astratte descrittive di tali episodi (ad esempio: OP; IC, ovvero, "obiettivo di possesso; intento cattivo" per descrivere episodi di vandalismo, ruberie, guerre di conquista, ma anche di litigi e scontri personali).

Sia i MOP che i TOP sono strutture modificabili a seconda dell'esperienza e permettono il filtraggio fra la MBT e la MLT. Come si vede, essi condividono ancora l'idea che la conoscenza possa essere rappresentata in forma eminentemente simbolica[28] e non poco devono all'impianto generale che soggiace alla Teoria del *Frame*, soprattutto nel modellizzare gli scenari in termini di aspettative che andranno attese o disattese. L'input, che costituisce nel modello di Schank la controparte di una MBT statica e rigida, è ancora una descrizione in linguaggio naturale, che viene compresa attraverso *cluster* di concetti predefiniti la cui modificabilità è relativa alle conseguenze disattese, più che ad una loro vera e propria scomposizione e ricomposizione. Appare evidente, perciò, che tale teoria è in grado di cogliere buona parte della complessità di una situazione intesa come episodio di azione, ma non affronta il problema della formazione concettuale e della categorizzazione. Di conseguenza, essa copre solo una parte dei processi analogici in senso lato descritti in precedenza. Eppure c'è un punto di contatto fondamentale con la TCCL e i modelli dei concetti fluidi, ed esso va ricercato nella struttura fondamentale di tali modelli.

---

[27] La definizione estesa che ne dà Schank è la seguente: «Un MOP consiste in un insieme di scene dirette verso il raggiungimento di un obiettivo. Un MOP è sempre una scena principale, il cui obiettivo è l'essenza e lo scopo degli eventi organizzati dal MOP» (Schank, 1982, p. 74). Come si vede, trattandosi di situazioni che descrivono sequenze di azioni, la definizione di MOP include come tratto fondamentale di essere *goal-oriented*.

[28] Questo è evidente anche nei due programmi di elaborazione dei testi che sono scaturiti dalle teorie di Schank sulla memoria dinamica: IPP e CYRUS. Il primo è un parser, un analizzatore grammaticale, parziale e integrato, capace di comprendere, nel senso dato a questo termine da Schank, attraverso l'aggiunta alla memoria episodica di informazioni specifiche e generalizzazioni a partire dall'elaborazione di testi giornalistici (una descrizione approfondita è data in Lebowitz, 1980). In esso è già operante il principio secondo cui la modificazione della memoria, attraverso la rievocazione e la percezione delle specificità della situazione in analisi, genera la comprensione. Tale principio è anche alla base di CYRUS (si veda Kolodner, 1981), un programma che riorganizza la propria memoria ogni volta che un nuovo fatto viene inserito attraverso E-MOP, cioè strutture di organizzazione della memoria episodica (*Episodic Memory Organization Packet*).

In precedenza si è visto che, affinché la TCCL possa funzionare, gli elementi base sono la MLT, la MBT e l'interazione fra queste due componenti. Un modello della teoria deve contenere non solo i due tipi di memoria, ma anche una parte elaborativa che li metta in corrispondenza. Dal punto di vista strutturale, dunque, il modello appare molto simile a quello proposto da Schank e può essere raffigurato con un'opportuna modifica, come nello schema della figura 2.4.

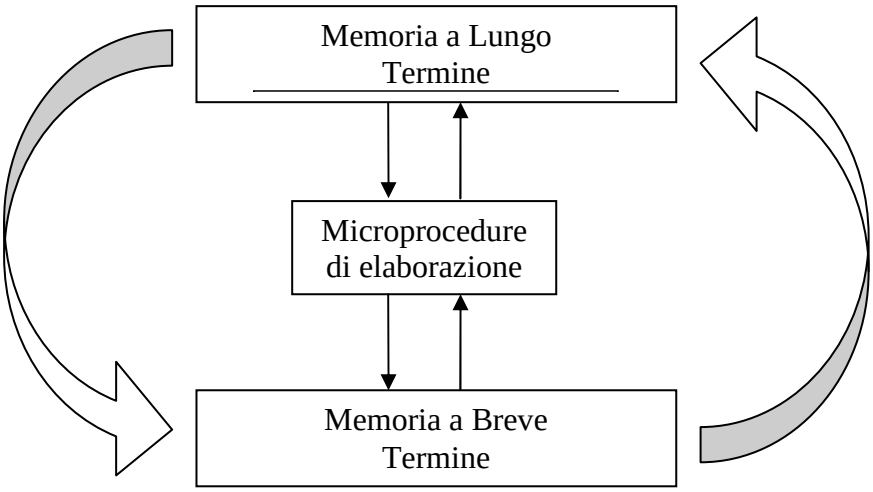

Fig. 2.4

La sostituzione del modulo MOP con un modulo costituito da microprocedure di elaborazione costituisce la vera differenza fra i due approcci e inaugura una differente visione dell'elaborazione della conoscenza per i sistemi intelligenti. È in questa mossa che va vista la principale innovazione nella modellizzazione dei processi cognitivi che riguarda i modelli dell'approccio subocognitivo e che può essere vista anche, per alcuni aspetti, come un preludio alla svolta connessionista nelle scienze cognitive. La mediazione fra MLT e MBT, e in generale fra tutte le forme di immagazzinamento dell'informazione nel sistema, avviene attraverso strutture di conoscenza atte a rendere il processo di rappresentazione della realtà – della situazione in oggetto – dinamico, contesto-dipendente e progressivamente adattivo sia alla conoscenza posseduta dal sistema, sia agli oggetti concreti di cui di volta in volta il sistema fa esperienza. In tal modo viene simulato quel processo di conoscenza continua che caratterizza il pensiero, enunciato dalla TCCL e raffigurato dalle due frecce laterali inversamente simmetriche della figura 2.4.

Per capire sia in cosa consista tale innovazione sia il meccanismo di funzionamento generale di questi modelli è opportuno risalire a due programmi che possono essere considerati diretti antecedenti di questo approccio: i modelli HEARSAY[29].

[29] Il riferimento ai modelli HEARSAY e in particolare a HEARSAY II come fonti ispiratrici dell'architettura subcognitiva è esplicito, ad esempio, in Hofstadter (1995e).

2.5.1 I modelli HEARSAY e la percezione distribuita del discorso

I modelli HEARSAY risalgono all'inizio degli anni Settanta e rappresentano uno dei tentativi, ancora nell'orbita dell'IA classica, di realizzare programmi in grado di comprendere il parlato. Essi furono realizzati all'interno del gruppo di ricerca diretto da Allen Newell alla Carnegie-Mellon University[30].

L'intuizione di fondo è quella di catturare il parlato, cioè di ricostruire il discorso che è stato pronunciato, attraverso programmi che facciano *ipotesi parziali* sugli enunciati espressi, cioè strutture sintattiche da riempire attraverso parole contenute in una base di dati lessicale. L'innovazione rispetto ai modelli precedenti risiede nell'utilizzo di sorgenti di conoscenza (*Knowledge Source*, KS) indipendenti, separabili le une dalle altre e contenenti informazione procedurale sulle operazioni da compiere. Al contempo, tali KS vengono poste in una relazione di *mutua non-interferenza* in modo da ottenere nel sistema un comportamento globale di cooperazione. Il meccanismo di cooperazione attua, attraverso le KS, un processo di produzione e verifica di ipotesi, cioè di creazione e valutazione, con riferimento alla base di dati generale dell'elaborazione del programma chiamata "lavagna". Ogni KS ha, dunque, la triplice funzione: di riconoscere il momento in cui, relativamente agli elementi presenti nella "lavagna", è in grado di contribuire positivamente al riconoscimento del particolare segmento di parlato su cui può operare; di formulare un'ipotesi (un enunciato parziale o un riempimento verbale); di valutare le ipotesi che sono già state avanzate da altre KS.

In particolare nel modello HEARSAY I (Reddy, Erman, Fennel, Neely, 1973) l'elaborazione del programma avviene ad un solo livello, quello delle parole, e le due attività fondamentali delle KS, la creazione di ipotesi (di proposizioni parziali da riempire con parole) e di valutazione delle ipotesi già formulate, hanno uno sviluppo, per così dire, orizzontale, cioè all'interno dello stesso livello del discorso. Ciò costituisce una forte limitazione, il cui superamento viene tentato con la realizzazione del modello più complesso HEARSAY II (Lesser, Fennell, Erman, Reddy, 1975), progettato per muoversi anche *verticalmente* fra i differenti livelli di comprensione del discorso, da quelli parametrico e fonetico a quelli frasale e concettuale: «Lo scopo principale della progettazione di HEARSAY II è quello di estendere i concetti sviluppati in HEARSAY I per la rappresentazione e la cooperazione della conoscenza al livello verbale a *tutti i livelli di conoscenza* necessari in un sistema di comprensione del parlato» (*ivi*, p. 13).

[30] Per una panoramica su queste ricerche si rimanda a Newell *et. al.* (1973).

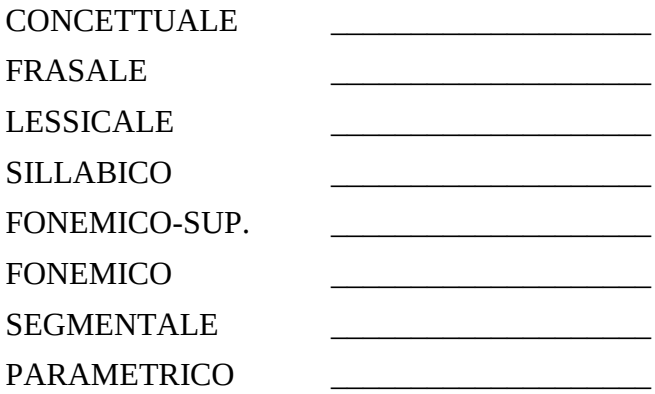

Fig. 2.5 (adattato da Lesser, Fennell, Erman, Reddy, 1975)

La figura 2.5 mostra la dimensione verticale del linguaggio in un'accezione generalmente condivisa dalla linguistica strutturale. Una delle linee guida ispiratrici di HEARSAY II è che le azioni effettuate ad ogni livello possono – e di fatto lo fanno in molti casi – influenzare le azioni compiute su altri livelli prossimi, o anche lontani, anche in due istanti di tempo consecutivi. Questo vuol dire che i livelli non sono vincolati in senso stretto dal punto di vista elaborativo e il programma, attraverso le sue KS, può compiere dei *salti di livello* (ad esempio, dal fonetico al lessicale o viceversa). Perciò, a partire dalla presentazione dell'input acustico, si passa attraverso un'opportuna descrizione del segnale, che produce una gamma di ipotesi di differenti segmentazioni in porzioni etichettate[31], per arrivare alla formulazione di ipotesi anche ai livelli più alti, come quello sintattico o concettuale. A tale processo *bottom up* si affianca un simultaneo processo *top down* di valutazione ed eventuale rivisitazione delle ipotesi già formulate. In tal modo, si crea una rete progressiva di rimandi fra livelli alti e bassi che guidano il programma alla resa in output di una verbalizzazione grafica quanto più fedele possibile all'input ascoltato.

Questa descrizione per sommi capi del programma mette in mostra alcuni suoi aspetti peculiari decisamente innovativi e rilevanti ai fini della nostra discussione dei modelli subcognitivi. Se ne possono individuare in particolare tre.

[31] La questione della segmentazione dell'input acustico linguistico è ancora oggi dibattuta a livello psicologico. In questa prospettiva si muovono, ad esempio, gli studi condotti da Jacques Mehler e dal suo gruppo di ricerca sull'acquisizione del linguaggio da parte dei neonati (per una panoramica su questi studi si veda Mehler, Dupoux (1990). Nel caso di HEARSAY II il programma si avvale di un apposito algoritmo di segmentazione basato su una collezione di parametri indipendenti. Per la descrizione di questo algoritmo si rimanda a Goldberg, Reddy, Suslick (1974).

In primo luogo, l'idea di un *parallelismo cooperativo* fra agenti di conoscenza[32]. Le KS sono strutture indipendenti fra loro, ma dipendenti dal contesto in cui agiscono. Esse compiono una funzione di creazione di ipotesi e di valutazione dei dati presenti nella "lavagna", un processo che riguarda sia le ipotesi effettuate in precedenza sia le materiale direttamente immesso dall'algoritmo di segmentazione dell'input percettivo. L'indipendenza delle KS è strettamente legata alla loro capacità di *azione locale*, mirata a quella parte della base di dati globale in cui *possono*, cioè *sono qualificate*, a intervenire. La struttura del discorso che ne deriva (il risultato dell'elaborazione generale del programma) si evolve fino a divenire sempre più stabile man mano che ipotesi a livelli diversi convergono verso una sorta di armonia globale. Perciò, ad esempio, ciò che da un parte viene generato dalla KS Ipotizzatore Fonemico dovrà essere alla fine in sintonia con ciò che produce la KS Ipotizzatore Semantico di Parole, senza tralasciare un accordo con la KS Parser Sintattico. Affinché questo processo possa avvenire è necessario che i diversi processi messi in atto dalle KS possano essere compresenti senza essere reciprocamente distruttivi grazie a opportuni meccanismi di controllo. Di conseguenza, dal punto di vista algoritmico le KS equivalgono a complesse regole di produzione simbolicamente formalizzate, in cui l'antecedente è costituito dalle precondizioni della possibilità della loro istanziazione e il conseguente dalle azioni prodotte dalle KS stesse.

HEARSAY II può dunque essere considerato un sistema a regole di produzione, e il parallelismo che mette in atto, implementato su un elaboratore sequenziale, consiste nel dispiegarsi indipendente delle KS a differenti istanti di tempo. Poiché non esiste, d'altra parte, un'unità di elaborazione centrale, ma è solo la soddisfazione delle precondizioni a determinare l'avvio della procedura di una KS, è opportuno che ogni volta che una KS valuta una sottoparte della situazione presente nella "lavagna" in vista delle creazione di nuove ipotesi, mantenga in memoria le condizioni da cui è partita al fine di un eventuale ripristino della situazione iniziale, ripristino necessario nel caso in cui il processo di valutazione evidenziasse il cambiamento di precondizioni di KS già istanziate in altre sottoparti della base di dati globale. Da questo punto di vista HEARSAY II potrebbe andare soggetto al problema dell'interferenza degli obiettivi tipico della pianificazione o anche all'anomalia di Sussman[33] se non venisse dotato di appositi meccanismi di bloccaggio delle azioni. A tali meccanismi si accosta, per contro, una sottocomponente algoritmica che ha lo scopo di

[32] Non a caso tali modelli sono tra i primi a essere sviluppati su calcolatori PDP-10 (*Programmed Data Processor model 10*), che sfruttano per la prima volta e in maniera estesa tutte le possibilità del *time sharing*. Tali calcolatori furono utilizzato per tutti gli anni settanta del Novecento nei laboratori di IA del MIT, a Stanford e alla Carnegie-Mellon University. Alcune funzioni del linguaggio assemblatore dei PDP-10 sono identiche a funzioni del linguaggio LISP, il linguaggio per eccellenza dell'IA. Questo suggerisce che l'utilizzo di particolari calcolatori, ovvero di particolari linguaggi per la codifica di algoritmi in forma di programma, non è del tutto irrilevante per il tipo di fenomeno che si intende simulare e, più in generale, per l'idea di *intelligenza* artificiale che si condivide. L'assunto funzionalista della realizzabilità multipla è una tesi metafisica la quale all'atto pratico della ricerca in IA non ha impedito che, in molti casi, si scendesse a compromessi con le esigenze e le possibilità effettive dei linguaggi di programmazione, al punto che si può sensatamente affermare che il funzionalismo racchiude in sé fin dalle sue origini, se non teoricamente almeno fattualmente, l'accettazione dei vincoli del substrato, con buona pace di chi vede in esso soltanto una riproposizione mascherata del dualismo (sia ontologico che epistemologico).

[33] Per una descrizione approfondita di tale problema e per alcune possibili soluzioni si rimanda a Sussman (1975).

valutare ogni ipotesi e di assegnarle un valore più o meno positivo per guidare la strategia globale nell'albero di ricerca, o, che è la stessa cosa, per rendere più probabile l'impiego di alcune KS a scapito di altre a seconda del valore dell'ipotesi cui sono correlate.

Il risultato finale è quello di una riproduzione del parlato (sull'interfaccia grafico) ottenuta senza la supervisione di un'unità di controllo globale, ma grazie all'impiego effettivo di una serie di differenti Sorgenti di Conoscenza che determinano il progressivo formarsi di collegamenti fra le ipotesi proposte, fino alla resa in output di un unico testo del discorso. È interessante notare come tale processo non corrisponda, pur riprendendone alcune caratteristiche, alla formulazione di un piano e alla sua scomposizione in sotto-obiettivi. Il *parallelismo cooperativo* consiste piuttosto nello sfruttare i suggerimenti che pezzi specifici di conoscenza possono apportare, parzialmente e indipendentemente, alla produzione del risultato finale *senza vincoli di implicazione sequenziale*, pur mantenendo una loro formulazione in forma logico-simbolica. Con le parole degli autori:

> L'approccio basato sulla decomposizione in sorgenti di conoscenza non è un tentativo di caratterizzare in qualche modo l'intero processo di riconoscimento e di applicare in seguito un'analisi di flusso di traffico alle sue elaborazioni interne al fine di decomporre il processo totale nelle KS interagenti a livello minimo. Piuttosto, le KS sono definite a partire da una qualche nozione intuitiva circa i vari pezzi di conoscenza che potrebbero essere incorporati in modo utile per aiutare il conseguimento della soluzione. (*ivi*, p. 17 in nota)

Un secondo aspetto importante del modello HEARSAY II è lo sfruttamento massiccio dell'interazione fra livelli *top down* e *bottom up*. Ad ogni livello del discorso vengono fatte diverse ipotesi, ad esempio lettere, se si tratta del livello fonemico di superficie, o parole, se si considera il livello lessicale, e così via. All'interno di ciascun livello, attraverso l'utilizzo di una grafo ad albero dotato di rami AND/OR, diverse opzioni vengono prese in considerazione fino al momento in cui alcune non diventino predominanti su altre. In particolare, ciò che risulta interessante è il fatto che il grafo ad albero, costruito su più livelli, ammette che i nodi non siano collegati soltanto in uno stretto rapporto padre-figlio/i, ma che i nodi figli possano avere più nodi padre nei livelli superiori. Il grafo, dunque, non può dirsi aciclico. La sua gerarchia rispecchia quella dei livelli in cui viene analizzato il linguaggio e non è strettamente interna e costitutiva del grafo in quanto struttura rigidamente gerarchica. Di conseguenza, il grafo è una *rete* piuttosto che un vero e proprio albero, i cui collegamenti esprimono funzioni di attivazione (o inibizione) bidirezionale, ma che mantiene tuttavia una differenziazione gerarchica fra gli elementi che la compongono (fonemi, sillabe, lettere, strutture sintattiche, frasi, ecc.).

Tutto ciò permette che ci sia una reciproca influenza fra livelli superiori e livelli inferiori, rispecchiando in parte le due idee sovrapposte che caratterizzano ogni visione della realtà, e degli specifici fenomeni del reale, per livelli: l'*emergenza* dei livelli superiori a partire da quelli inferiori

e la *dipendenza dal contesto* da parte dei livelli inferiori, laddove il contesto è costituito da quelli superiori. All'interno di un sistema caratterizzato da un passaggio bidirezionale di informazione le relazioni *top down* e *bottom up*, tendono a diventare simmetriche, e, di conseguenza, l'albero ramificato una rete di relazioni, che arriva ad assumere il ruolo di *meta-contesto* globale dell'elaborazione. Nondimeno, il fatto che sia ancora necessaria una distinzione fra *livelli bassi* e *livelli alti* è dovuto alla presenza di un input costituito da materiale percettivo. In altri termini, *la percezione sembra implicare in ogni caso un certo quantitativo di gerarchia nel processo conoscitivo*.

Infine, il modello HEARSAY II è un modello che anticipa in parte le tendenze dell'IA degli anni Ottoanta attraverso la simulazione di un fenomeno percettivo come la comprensione del parlato. A differenza dei programmi quasi coevi sviluppati a Yale dal gruppo di ricerca guidato da Roger Schank, nel caso di HEARSAY II non si tratta di attuare una comprensione profonda del significato del discorso, inteso come narrazione, ma solo di comprendere il discorso come atto di espressione articolata foneticamente. Per fare questo il programma si avvale di un'architettura mirante a cogliere gli aspetti percettivi di basso livello e quelli di alto livello, e sfrutta una concezione strutturale del linguaggio, come gerarchia di livelli costituiti da elementi composizionali ma in una reciproca relazione di interdipendenza. Tale modello è stato, perciò, anticipatore di tutte quelle tendenze dell'IA che non considerano aspetto percettivo e aspetto cognitivo come due componenti distinte e distaccate, ma come due segmenti distinti e correlati dello stesso processo. Ciò costituisce uno degli assunti principali delle scienze cognitive degli ultimi decenni, per le quali, come a volte sottolineano con un'enfasi a discapito di altre caratteristiche, l'essere calati all'interno dell'ambiente percettivo (*situatedness*) è uno dei tratti caratteristici dei sistemi intelligenti naturali e artificiali.

2.5.2 La scansione parallela a schiera

La nascita di un sistema che sfrutta il parallelismo cooperativo a Carnegie-Mellon nel gruppo di ricerca diretto da Allen Newell potrebbe suggerire che, nello spirito dell'IA di Newell e Simon (Newell, Simon, 1972) dedito alla simulazione dei processi mentali e non solo alla riproduzione dei loro risultati, cioè in accordo col quale il *meccanismo* del pensiero assume una posizione preminente rispetto alla *prestazione* cognitiva, i modelli HEARSAY siano un altro tassello di questa impostazione teorica della ricerca simulativa. In realtà, pur non disconoscendo i pregi di questi modelli, va tenuto conto che accanto all'immissione di una certa misura di probabilismo a regolare l'entrata in scena delle sorgenti di conoscenza non si trova, da parte degli autori dei modelli stessi, l'affermazione che i processi di pensiero umano seguano un andamento di questo tipo. Infatti, le KS sono, a conti fatti, soltanto porzioni di programma che «posseggono la capacità processuale in grado di risolvere alcuni sottoproblemi, date le appropriate circostanze della loro attivazione» (*ivi*, p. 16). La loro peculiarità, specifica di questo approccio all'IA, consiste nel fatto che le KS predispongono

la possibilità di *effettuare piani il cui andamento e il cui esito non solo non sono prevedibili all'inizio, ma non lo sono neanche in un punto intermedio dell'elaborazione*, rispecchiando in questo la reale pratica cognitiva umana di essere sempre dinamicamente dipendente da un contesto *nell'esercizio delle sue funzioni*.

Si è affermato in precedenza che la peculiarità di HEARSAY II risiede nel modo in cui il parallelismo viene utilizzato. Le KS hanno precondizioni che, se soddisfatte, giustificano la loro esecuzione. Tali precondizioni sono, in genere, disposte su un doppio livello. Infatti, le precondizioni dell'azione creativa (di ipotesi) sono soddisfatte dall'esecuzione della parte valutativa della KS, valutazione che ha precondizioni che devono a loro volta essere soddisfatte dalla presenza di specifici elementi nella lavagna, affinché la procedura espletata dalla KS venga chiamata. In sintesi, l'intero processo, che può essere visto, ricordiamolo, come un sistema a regole di produzione, è basato sulla valutazione iniziale della base di dati globale, cioè la "lavagna", che esamina le precondizioni delle precondizioni di ogni KS, assegnando loro un valore di rilevanza ai fini della presentazione di una possibile ipotesi completa finale. Questo processo ha lo scopo di evitare il dispendio di risorse computazionali con l'esecuzione completa di numerose KS e di far sì che solo le più promettenti vengano messe in pratica, lasciando non eseguite quelle le cui precondizioni sono valutate più basse rispetto a un valore di soglia. L'istanziazione di KS in diverse parti localizzate della base di dati avviene sequenzialmente, ma in maniera asincrona, nel senso che, come si è visto, posti alcuni meccanismi di non interferenza reciproca, le KS agiscono in maniera indipendente incastonandosi nell'esecuzione del programma, ad intervalli di tempo diversi, del processo globale di valutazione della situazione della base di dati.

L'*aspetto parallelistico ma anche gerarchico* dell'elaborazione di HEARSAY II ha ispirato la terza caratteristica dei modelli subcognitivi: la *scansione parallela a schiera*. Tale dicitura sta a significare un processo di elaborazione che a partire da una condizione di estremo parallelismo va verso l'individuazione di cammini sempre più promettenti man mano che l'elaborazione avanza. Essa può essere considerata una strategia di ricerca in profondità non deterministica e direttamente guidata dal materiale presente nella base di dati. Al primo stadio tutte le possibilità vengono prese in considerazione. Al secondo stadio, vengono considerati, sempre in maniera parallela (asincrona), solo i nodi che hanno ricevuto una valutazione maggiore. Al terzo c'è ancora un raffinamento, e così via. Anche in questo caso, il processo è compiuto da microprocedure esplorativo-valutative, che possono essere costruttive o distruttive.

È facile ora vedere come tale processo si adatti all'architettura descritta nella figura 2.4 e come gli elementi di HEARSAY II trovino una corrispondenza che svela anche il ruolo del modulo elaborativo centrale. La base di dati globale, la "lavagna", corrisponde alla Memoria a Breve Termine. Le KS corrispondono alle microprocedure elaborative, chiamate *codelets* ("codicelli"). Manca una corrispondenza fra gli elementi di HEARSAY II e la Memoria a Lungo Termine dei modelli subcognitivi. La MLT è di fatto una rete semantica, la cui attivazione rispecchia in senso

astratto l'attività delle microprocedure nella MBT. D'altra parte, la funzione delle microprocedure è la stessa in entrambi i tipi di modelli. Esse incorporano una quantità di conoscenza che guida l'elaborazione verso la formazione di ipotesi sempre più raffinate e complete. Da ciò discende che quello che in precedenza è stato chiamato modulo elaborativo centrale non è un'unità di controllo del processo di elaborazione, ma una semplice lista di operazioni legata a un algoritmo probabilistico, la cui esecuzione porta alla costruzione di strutture, raggruppamenti e collegamenti sempre più complessi all'interno della MBT.

Anche nel caso dei modelli subcognitivi le microprocedure sono indicizzate con un valore che ne indica la rilevanza ai fini dell'esecuzione. Tale valore è chiamato *urgenza*, perché tanto più è alto, tanto prima verrà chiamata ad agire la microprocedura cui è assegnato. Esso dipende dalla valutazione della situazione in corso, nel senso che ogni microprocedura che viene eseguita prima di terminare la sua funzione decide quale valore assegnare ad una sua copia nella lista delle microprocedure: più la strada che prende sembra promettente, più la sua discendenza riceve un valore alto. Tuttavia, tale valore è determinato anche dalla quantità di attivazione dei nodi della rete semantica, che sono collegati alle loro specifiche microprocedure, e da un'altra variabile, la *temperatura*, che indica lo stato generale del sistema. Più la soluzione sembra vicina, più la temperatura si abbassa e più le microprocedure ricevono valori alti se sono una continuazione dei percorsi già intrapresi dal programma. Se si verifica una fase di stallo, la temperatura si alza e le urgenze delle microprocedure vengono livellate affinché il processo ricominci con una forte dose di parallelismo[34].

[34] La funzione della variabile temperatura in questi modelli non va confusa con quella più tipica dei modelli connessionisti, con la quale ha delle parziali parentele. Per questi, ad esempio con riferimento alla macchina di Boltzmann (Ackley, Hinton, Sejonowski (1985), Hinton, Sejnowski (1986)), si parla di temperatura come di una certa quantità di energia che viene aggiunta ai nodi della rete per provocare una maggiore oscillazione nei valori di attivazione. Essa determina, in altri termini, la non linearità della rete in misura proporzionale alla quantità di energia che viene immessa nei nodi. Durante l'apprendimento della rete essa viene progressivamente ridotta e si parla di "raffreddamento simulato" (*simulated annealing*), cioè il sistema viene ricondotto a un andamento più lineare affinché possa giungere a uno stato di equilibrio, che corrisponde alla soluzione o a una delle soluzioni possibili. Nel caso delle reti connessionistiche la diminuzione della variabile temperatura (raffreddamento) è controllata dall'"esterno", cioè da una procedura appositamente *pre*-programmata e indipendente dall'andamento della rete (viene fatto in genere un paragone con il modo in cui determinati metalli vengono raffreddati dopo il processo di fusione per evitare il formarsi di strutture impure). Nei modelli subcognitivi la variazione della temperatura è strettamente dipendente dal processo generale di elaborazione e può oscillare anche più volte fra aumenti e diminuzioni. L'analogia, in questo secondo caso, è con la biologia degli organismi. A una maggiore attività metabolica corrisponde un aumento di temperatura e, viceversa, un aumento della temperatura indica l'accelerazione dell'attività metabolica.

Per trovare un precedente nei sistemi di IA tradizionali, cioè simbolici e basati sul calcolo dei predicati, della funzione di auto-monitoraggio svolta dalla temperatura si può forse fare riferimento al programma di simulazione dei processi nevrotici sviluppato da Colby nel corso degli anni sessanta (Colby, 1963). In esso, alcune subroutine misurano il livello di pericolo, di eccitazione, di piacere, di autostima e di benessere del programma; in altri termini, la loro funzione è quella di esternare le componenti emotive di esso. Tuttavia, va sottolineato che il programma lavora su sistemi di credenze e i risultati quantitativi prodotti dalle subroutine derivano direttamente dalla conoscenza esplicitamente rappresentata dal programmatore in una serie di matrici in cui a termini del linguaggio naturale sono associati valori numerici. Si può, dunque, affermare che il programma di Colby *simula* il proprio stato emotivo in conformità a quello di un essere umano nevrotico nell'elaborare, in modo puramente sintattico, un determinato insieme di credenze più o meno conflittuali, mentre nei modelli subcognitivi la variabile temperatura non ha tali pretese simulative, espletando solamente una funzione di auto-controllo sull'andamento stocastico del programma. Nel primo caso il programma *simula* l'auto-valutazione, nel secondo *stima effettivamente* la propria elaborazione. Tuttavia, non

Nel procedere dell'elaborazione, perciò, si ha una generale tendenza di avanzamento dallo stocastico al deterministico. Se più percorsi all'inizio sembrano promettenti, perché all'interno di una situazione (nella MBT) più elementi possono essere correlati attraverso differenti aspetti (due oggetti uguali o con la medesima funzione o con la stessa relazione spaziale), al momento in cui le correlazioni saranno trasformate in raggruppamenti stabili soltanto quelle che permettono ulteriori livelli di correlazione, cioè correlazioni a un livello più astratto, verranno portate avanti dall'elaborazione. Tutto ciò avviene senza che ci sia nessun tipo di unità di controllo centrale, ma solo grazie alla *selezione* dei percorsi più promettenti, e quindi all'*adattamento* delle conoscenze pratiche, cioè procedurali, possedute dal programma alla situazione presa in esame.

Il fatto di procedere per livelli di raggruppamento sempre maggiore, o per livelli di correlazione fra elementi, in base ad una qualche relazione specifica (identità, successione, ecc.) *strettamente dipendente dal contesto*, non vincola il programma a un passaggio automatico in avanti o all'indietro fra i livelli. Piuttosto i raggruppamenti, cioè le relazioni categoriali fra gli elementi della situazione in esame, sono compiuti, come era già in HEARSAY II, a livello locale e in maniera asincrona, cosicché elementi diversi possono essere collegati in modi diversi fra loro in tempi diversi, ma senza che le operazioni compiute su alcuni elementi influenzino necessariamente tutte le altre operazioni compiute nello spazio di lavoro (la MBT). I differenti livelli di astrazione rappresentano, dunque, il *passaggio intensionale*, effettuabile nei due versi *bottom up* e *top down*, fra occorrenza (*token*) dei concetti e tipi (*type*) dei concetti. Inoltre, l'elaborazione parallela, guidata dall'attivazione dei nodi concettuali nella rete semantica, garantisce di ritorno che ogni nodo possa essere di volta in volta considerato occorrenza o tipo, all'interno dei legami categoriali complessivi che instaura con gli altri nodi, e a seconda di ciò che *conviene* a quel punto dell'elaborazione e in quel particolare aspetto locale della situazione globale presa in esame. Il fatto che un concetto non sia rigidamente fissato come *token* o come *type* rispetto agli altri concetti permette la costruzione di gerarchie di relazioni (ad esempio, si può avere una "successione di identità", ma anche un'"identità di successioni") che nella logica dei predicati sono *esprimibili in maniera diretta*, cioè molto simile al modo in cui lo fa il linguaggio naturale, a partire dal calcolo predicativo del secondo ordine in su, in cui si ha la possibilità di quantificare su relazioni e proprietà.

Inoltre, a differenza che nel modello HEARSAY II, dove i livelli (in quel caso, del linguaggio parlato) sono fissati in maniera predeterminata nella struttura del programma, nei modelli subcognitivi la creazione di livelli di astrazione è un fatto intrinseco al procedere dell'elaborazione, senza limiti predeterminati. In tal modo, si vuole catturare la capacità potenzialmente illimitata dell'applicazione ricorsiva dei concetti nel creare strutture concettuali, definite "scheletri concettuali", sempre più complesse, senza che il processo sia determinato da altro che dalla conoscenza che viene di volta in volta, nel corso di *ogn*i elaborazione, messa in atto dal programma

---

sembra inopportuno vedere fra i due casi un legame, seppure, verosimilmente, soltanto da un punto di vista euristico. Per una descrizione e una discussione del programma di Colby si rimanda a Boden (1986).

con il gioco di rimandi fra MLT (rete semantica) e MBT (spazio di lavoro) attraverso l'applicazione delle microprocedure. Si vedrà in seguito come ogni modello sfrutta caratteristiche più o meno diverse di questo schema generale.

Occorre ancora dire che, a differenza di quello che era il MOP nel modello teorico della memoria dinamica di Schank, cioè una struttura per il *packaging* e l'*unpackaging* dei dati, il modulo algoritmico delle microprocedure si differenzia nell'essere, più che un insieme di strutture preconfezionate, un modulo funzionale di *creazione di strutture*, la rappresentazione delle quali va cercata nei livelli di attivazione della rete semantica, in quanto scheletro concettuale, e nello spazio di lavoro, come collezione di agglomerati fra gli elementi presenti. Per marcare questa distinzione, la figura 2.4 va a questo punto perfezionata. Il modulo microprocedurale, affinché possa dar vita alle due frecce grandi che raffigurano il *loop* cognitivo, deve essere considerato come *modulo di mediazione parallelistica*, il cui intervento nell'elaborazione è in ogni momento potenzialmente differenziato, pur con i vincoli imposti dalla tendenza deterministica del processo[35]. Dunque, lo schema generale dell'architettura dei modelli descritto in precedenza (fig. 2.4) può essere modificato come in figura 2.6.

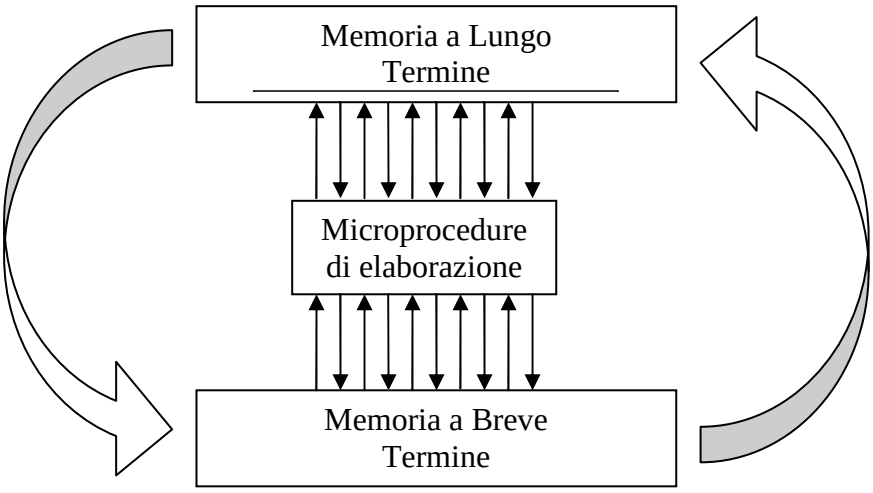

Fig. 2.6

L'idea guida dei modelli subcognitivi, perciò, è che la rappresentazione della conoscenza è funzione emergente dell'elaborazione e si dà in diverse forme, anche se correlate, nelle differenti parti dell'architettura computazionale. *La rappresentazione della conoscenza è, in altri termini, funzione del suo uso* e la distinzione, non elaborativa ma descrittiva, fra rappresentazione e uso della conoscenza costituisce uno degli aspetti principali e più innovativi di questo approccio all'IA. Ciò che rende possibile l'implementazione di tale idea è il *parallelismo procedurale* che si

[35] La stabilizzazione del processo elaborativo garantisce l'arrivo a una soluzione. Attraverso la modulazione del valore della temperatura si possono stabilire misure diverse di "quantità stocastiche" nell'elaborazione. Questo è uno dei modi in cui si può variare lo stile individuale del programma.

contrappone al *parallelismo rappresentazionale*, cioè di codifica della rappresentazione, tipico dell'approccio connessionista.

Una delle analogie ricorrenti per spiegare in che modo agisce il meccanismo della scansione parallela a schiera è quella con il metabolismo cellulare (Hofsdtadter *et. al.*, 1995, Mitchell, 1993). Nel citoplasma della cellula, senza la direzione di un'unità di controllo centrale, ma grazie all'interazione di migliaia di processi enzimatici, vengono costruiti tutti i tipi di molecole necessarie al funzionamento interno della cellula stessa (e in alcuni casi specifici anche esterno). Tali molecole occupano uno spettro di complessità crescente. Quelle più semplici, come ad esempio l'H_2O o il CO_2 (l'acqua e il biossido di carbonio), sono costituite da legami molto forti e stabili. Quelle più complesse, gli amminoacidi, le catene di amminoacidi e le proteine, sono costruite a partire dalle molecole più semplici tenute insieme tra loro da legami meno forti e stabili. I processi di costruzione coinvolgono una serie di semplici operazioni compiute dagli enzimi, operazioni disposte in una sequenza predefinita, anche se non sono compiute tutte da uno stesso enzima. Perciò, ogni passo della costruzione di una molecola dipende dal fatto che i precedenti siano stati effettuati e che il contesto sia adatto, tanto quanto dipende dalla macchina enzimatica che lo mette in atto. Solo la presenza di materiale pronto ad entrare in un qualche specifico stadio reattivo fa sì che tale stadio si inneschi. Analogamente, è la presenza eccessiva di certe sostanze nella cellula a far generare enzimi, cioè altre molecole, che inibiscano gli enzimi di costruzione delle sostanze in eccesso. *Tutto il processo si autoregola, ma non autocrea*, a partire dal materiale disponibile, seguendo sequenze di operazioni codificate a monte nella sequenza genetica che esprime l'informazione necessaria per la formazione delle macchine enzimatiche.

L'analogia con la cellula suggerisce almeno tre cose: che la costruzione di strutture superiori, più complesse e di alto livello, deve essere *regolata* più che *guidata*; che essa deve avvenire a partire da elementi più semplici, il cui legame è più robusto, cioè più immediato e meno soggetto ad ambiguità (due o più atomi tendono a legarsi molto più facilmente nel legame più forte che possono costituire; due o più molecole sono maggiormente dipendenti dal contesto chimico in cui si trovano nel generare questa o quella reazione specifica); che, di conseguenza, al di là di una netta differenziazione del livello atomico rispetto a quelli superiori, l'ambiente in cui la formazione di strutture avviene deve essere dotato di una *generale omogeneità*, la quale permette tanto la costruzione quanto la distruzione di strutture diverse a partire dagli stessi elementi (cioè, elementi dello *stesso tipo*, sotto un qualche aspetto).

L'analogia con il metabolismo cellulare potrebbe anche sembrare un suggerimento dell'idea che i processi mentali siano biologicamente afferrabili attraverso la loro riduzione all'interazione fra le cellule (neuronali) su cui avvengono. Tuttavia, non va intesa in questo senso. L'utilizzo a fini esplicativi dell'attività del citoplasma è solo metaforico. Altre metafore potrebbero andare ugualmente bene e riguardare pratiche e comportamenti sociali, come ad esempio, il complicato procedimento casuale e selettivo di scelta di un partner o la suddivisione di fondi all'interno di un

istituto di ricerca attraverso un meccanismo di coagulazione delle risorse intorno ai progetti più interessanti secondo criteri che, come è noto, spesso si muovono su uno spettro che comprende ragioni scientifiche, etiche e sociologiche. Appare chiaro che la forza di queste analogie sta nell'indicare somiglianze strutturali e funzionali, che diventano nei modelli punti di convergenza dell'architettura in base a cui sono costruiti.

Se, dunque, non è possibile considerare l'analogia con l'attività cellulare come indice di una plausibilità biologica *forte* di questi modelli[36], nondimeno per essi viene rivendicata una plausibilità psicologica che è strettamente correlata alla loro architettura imperniata sul parallelismo procedurale e sull'andamento stocastico convergente a stati determinati e univoci. Si presume che ciò rispecchi l'attività inconscia della mente, che consciamente invece non sfugge alla sequenzialità dell'attenzione cosciente. Il seguente passo è illuminate al proposito, riassumendo la posizione teorica generale che supporta i modelli *subcognitivi*:

> Il punto di vista effettivo del sistema si sviluppa nel tempo in questo modo: si esplora in continuazione un "alone" probabilistico di molte direzioni *potenziali*, le più promettenti delle quali tendono a divenire *effettive*. Questo aspetto [dei modelli], per inciso, riflette il fatto, importante dal punto di vista psicologico, che l'esperienza conscia è essenzialmente unitaria, anche se risulta, come è ovvio, da molti processi paralleli inconsci. (Mitchell, Hofstadter, 1994, p. 248)

Gli elementi "psicologici" presenti in questa dichiarazione di principio sono molteplici: la dinamica temporale del pensiero, la sua manifestazione conscia, il suo agire inconscio come somma equilibrata di molteplici interazioni – non necessariamente quella del substrato neurale, ma situata prevalentemente *ad un livello superiore*, subcognitivo appunto. Tuttavia, va sottolineato che la derivazione dell'esperienza cosciente da «molti processi paralleli inconsci» non è così ovvia come sembra a prima vista, almeno per quanto riguarda il modo in cui essa "risulta" da essi.

Quale è lo scopo generale, dunque, delle architetture subcognitive e delle loro traduzioni in algoritmi e programmi? Che cosa ci dobbiamo aspettare che spieghino?

In primo luogo, la *teoria dei concetti* che esse mettono in gioco e in secondo luogo la *teoria del ragionamento* che è implicata nella loro costruzione e che rimanda a una precisa metafisica del sistema mente-cervello. Tuttavia, un'analisi del loro funzionamento specifico, affrontata nei prossimi capitoli, oltre a produrre un qualche tipo di risposta alle domande appena formulate, potrà produrre come risultato epistemologico un'ulteriore chiarificazione di come teoria e pratiche nelle scienze cognitive abbiano un legame peculiare e diverso da quello delle altre scienze. Dovrebbe essere manifesto alla fine di questo percorso che le ricerche sui processi del pensiero *a un certo*

[36] D'altra parte, la questione della plausibilità biologica di questi modelli c'è e riguarda, naturalmente, le questioni ontologiche relative alla metafisica che supporta la teoria dei processi mentali di cui i modelli cognitivi della subcognizione sono implementazione. Torneremo su questo argomento ancora una volta nel capitolo conclusivo, dopo aver discusso nel dettaglio i modelli più significativi.

livello non possono fare a meno di una componente simbolica coinvolgente *una qualche accezione di rappresentazione* e presente necessariamente nei compiti analogico-percettivi e semantici, che i modelli sono chiamati ad affrontare e, così facendo, a spiegare.

Capitolo 3

I MODELLI SUBCOGNITIVI DELLA PERCEZIONE ANALOGICA

3.1 Una possibile classificazione

In questo capitolo verranno esposti e commentati diversi modelli della percezione sorti all'interno dell'approccio subcognitivo allo studio dei processi di pensiero. Si deve tener presente che la percezione di cui si parla è quella che abbiamo definito "di alto livello", cioè fortemente intessuta di apporto concettuale e categoriale, e intesa come motore cognitivo dei differenti processi analogici di cui si è discusso nel precedente capitolo.

I modelli presi in considerazione sono frutto di più di venti anni di ricerche ad opera di Hofstadter e collaboratori, il così detto *Fluid Analogies Research Group* (FARG). Seppur con metodologie di lavoro diverse e privilegiando a volte alcuni e a volte altri aspetti del processo percettivo-analogico, essi possono essere tutti ascrivibili ai principi esposti nel capitolo precedente.

Di tali modelli si possono dare differenti esposizioni e classificazioni. Quella più ovvia è senz'altro di considerarli in una prospettiva storica attraverso la quale constatare l'apporto specifico di ognuno di essi nei differenti periodi che hanno attraversato la scienza cognitiva negli ultimi tre decenni. Tuttavia, il prezzo da pagare per questa scelta è quello di sacrificare in maniera eccessiva i rimandi interni ai diversi modelli e il processo evolutivo che in determinati casi lega alcuni di loro in modo più stretto dal punto di vista dei fenomeni indagati e dei principi in gioco nella loro progettazione. Un'altra possibile classificazione consiste nel loro raggruppamento in due macroaree, quella dei modelli che più specificamente si occupano della percezione di alto livello e quella dei modelli volti alla creazione di analogie. Anche in questo caso, però, la distinzione non è netta e spesso i due obiettivi (simulativi) risultano intrecciati in maniera inseparabile, anche se la prestazione del programma può, a causa del dominio in cui agisce o della sua interfaccia grafica o del grado effettivo di realizzazione, mettere in luce un compito piuttosto che l'altro, pur presente nel processo messo in atto dal modello e *descritto* dall'architettura del modello.

Si è scelto, perciò, in questo capitolo di presentare i modelli accostando quelli che operano nello stesso dominio, o in domini molto simili, e costituiscono evoluzioni successive nel tentativo di affrontare il medesimo problema. Per quanto l'architettura di fondo dei modelli sia in qualche

misura sempre basata sulle stesse componenti teoriche strutturali, sono riscontrabili differenze che mostrano in alcuni casi una differente impostazione nel dare una risposta ai problemi affrontati e che evidenziano alcuni aspetti a scapito di altri. Il dominio problemico per cui i modelli sono costruiti contribuisce a questa differenziazione, richiamando diversi elementi in gioco nei processi (di pensiero) attuati per affrontare i compiti prescelti. Questa impostazione non si concretizza, né dovrebbe farlo, in proposte *ad hoc* per l'architettura dei modelli, se non per aspetti superficiali, che riguardano l'interfaccia, o di livello più alto rispetto a quello dell'elaborazione principale del programma che costituiscono un arricchimento del modello e non uno stravolgimento dei vincoli strutturali su cui è costruito. Se non fosse così, da una parte diverrebbe difficile *valutare l'efficacia esplicativa* dei modelli, dall'altra essi perderebbero valore nell'ottica di una valutazione globale, derivabile dalla loro comparazione, di questo approccio allo studio dei processi cognitivi. L'utilizzo di differenti domini, infatti, permette di ottenere risultati diversi a partire da medesime premesse, cioè dagli stessi vincoli, in vista di una generalizzazione dei principi teorici implicati nell'indagine simulativa.

Quale è lo scopo di questa operazione? Innanzitutto, quello di procedere alla ricostruzione e alla valutazione dei principi messi in atto nella progettazione e nello sviluppo di questi modelli, che incorporano, cercando di esserne realizzazione pratica, alcune determinate teorie in merito ai problemi della conoscenza, della concettualizzazione e dei meccanismi attraverso cui si attua il ragionamento, o perlomeno, alcune forme di ragionamento, quello analogico in particolare. Alla fine di questa disamina dovremmo essere in grado di scorgere il filo unitario che lega queste ricerche e di avanzare ipotesi circa l'efficacia di questo tipo di ricerca all'interno dell'orizzonte delle scienze simulative in generale e della filosofia della mente che le supporta. Le domande che ci interessano, infatti, sono relative alla misura in cui i risultati attesi sono stati conseguiti e alle riflessioni critiche che se ne possono trarre: che tipo di percezione effettivamente mettono in opera questi modelli? È plausibile, alla luce di questi modelli, l'impianto di principi che regolano quello che abbiamo chiamato approccio subcognitivo all'IA? Se si vuole, la domanda finale posta nei termini più generali è la seguente: che cosa dimostrano questi modelli simulativi?

Inoltre, se si accetta l'assunto che per fare filosofia della scienza, di una determinata scienza, occorre conoscere i risultati raggiunti, ciò è tanto più vero nel campo dell'intelligenza artificiale e delle scienze cognitive, che affrontano argomenti profondamente connessi con la riflessione filosofica tradizionale, la quale non può non esserne influenzata e, allo stesso tempo, determinarne in parte le prospettive di indagine globali, ma anche le ricerche specifiche. Questo ci introdurrà alle considerazioni finali di questa dissertazione, sviluppate nel prossimo capitolo.

3.2 La proposta di un modello teorico

Dovendo scegliere un punto di partenza, sembra opportuno rintracciarlo nell'antecedente più diretto di questi modelli, delineato come abbozzo teorico di sistema cognitivo da Hofstadter nel suo celebre volume *Gödel, Escher, Bach* (Hofstadter, 1979). L'influenza che questo libro ha avuto sulla cultura contemporanea e nel complesso degli studi sulla mente è molto vasta e a prima vista quasi indecifrabile, tanto quanto poteva essere imprevedibile – e non prevista – prima della sua uscita[1].

Il capitolo diciannovesimo di *Gödel, Escher, Bach* è dedicato alle prospettive future dell'IA, che in quegli stessi anni attraversava una fase di crisi e di cambiamento dovuta al palesamento di una serie di problemi relativi alla conoscenza e alla dotazione epistemica che un sistema intelligente deve possedere per potersi definire tale e perché la sua *azione* sia giudicabile, a ragione, "intelligente" secondo i canoni del pensiero umano.

Tale problema portò a una serie di risultati importanti sia per quanto riguarda lo sviluppo di nuove forme di memoria e di immagazzinamento dei dati in un programma (si pensi alle reti semantiche, ai *frame*, agli *script*, e così via), sia dal punto di vista della riflessione filosofica che si occupava, a quel tempo, di argomenti correlati. Si può affermare che proprio in quegli anni le strategie simulative dei processi di pensiero, con i loro risultati pratici, divengono uno dei principali interlocutori nelle controversie sulla natura "semantica" del pensiero e sul problema della rappresentazione, che è come dire, della memoria, dei concetti e delle idee, temi chiave della riflessione gnoseologica ed epistemologica da tempi molto più remoti della nascita della nozione di IA e del suo affermarsi come disciplina consolidata, al tempo stesso problematica e riconosciuta.

Nell'affrontare il problema di quali caratteristiche siano necessarie a un sistema di IA per esibire capacità intelligenti, Hofstadter propone un modello teorico di programma, a partire dall'individuazione di un dominio adatto alla sperimentazione di capacità percettive e concettuali tipiche dell'uomo: il dominio dei problemi di Bongard (Bongard, 1970). Questi sono problemi di riconoscimento di forme («*patterns*»), nei quali a un soggetto vengono sottoposti dodici riquadri raffiguranti forme geometriche di vario tipo e divisi in due gruppi, uno di destra e uno di sinistra (fig. 3.1). Lo scopo è quello di trovare in che modo, cioè secondo quale *proprietà comune*, i sei riquadri di destra differiscono da quelli di sinistra. Ad esempio, si può dare il caso che nei primi sei riquadri ci sia una prevalenza di cerchi dentro triangoli e nei secondi sei ci siano, invece, molti triangoli dentro cerchi. Esistono anche problemi in cui la forma delle figure all'interno è indifferente e ciò che conta è, magari, il loro essere raggruppate o sparpagliate. Da questi esempi si comprende che la soluzione dei problemi proposti da Bongard non è dovuta a una conoscenza molto approfondita della geometria, bensì piuttosto alla capacità di enucleare analogie a un certo livello di

[1] Soltanto la ricostruzione degli influssi avuti sugli studiosi di differenti discipline dalla sua uscita ad oggi potrebbe costituire argomento per un volume di storia delle idee, se i tempi non fossero ancora troppo prematuri per questo tipo di indagine. Il volume è stato pubblicato per la prima volta nel 1979 in edizione americana e tradotto in molte lingue, tra cui anche il russo e il cinese. La prima edizione italiana è del 1984.

astrazione concettuale fra i gruppi di riquadri e di metterle, poi, a confronto. Si tratta, in altri termini, di un doppio compito analogico, la cui soluzione consiste nel trovare la giusta *relazione meta-analogica* fra i due insiemi di figure.

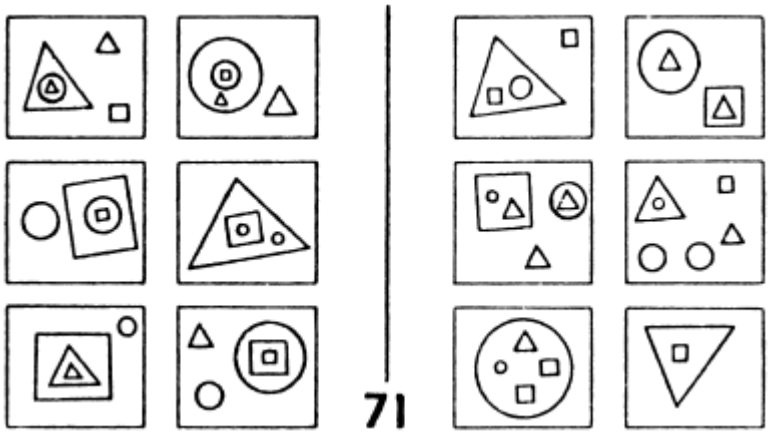

Fig. 3.1 - Problema di Bongard n. 71 (tratto da Bongard, 1970)

Hofstadter individua alcuni tratti essenziali di un programma in grado di risolvere questo tipo di compiti, che fondono allo stesso tempo capacità percettive e concettuali. Appare chiaro, infatti, che solo attraverso un uso opportuno delle *descrizioni* che il programma fa della situazione in oggetto è possibile arrivare a una soluzione del problema di Bongard. La questione delle descrizioni è fondamentale da più punti di vista e si riallaccia ai *frame* in quanto tecnica di rappresentazione della conoscenza, introdotta negli anni settanta da Marvin Minsky (1975)[2]. I *frame* vengono definiti da Hofstadter come «una *rappresentazione algoritmica del significato*» (Hofstadter, 1979, p. 697) e determinano, a loro volta, la struttura dei concetti che di essi fanno parte, poiché «i concetti vengono compressi e distorti dai contesti nei quali sono inseriti a forza» (*Ibidem*). Perciò, il problema di come dare descrizioni affidabili e pertinenti di una situazione risulta inscindibile da quello dei concetti che vengono impiegati nella descrizione. L'obiettivo, per quanto riguarda i Problemi di Bongard, è quello di arrivare ad una rappresentazione dei due insiemi di riquadri che

[2] È noto che i *frame* sono, in termini generali, schemi attraverso cui l'informazione viene strutturata a partire da un nucleo comune condiviso da tutte le situazioni e gli oggetti che possono essere descritti attraverso lo stesso *frame*. Essi sono dotati di terminali (*slot*) da riempire con le caratteristiche specifiche della situazione in oggetto, quale può essere, ad esempio, un particolare esempio del concetto STANZA o un particolare esempio del concetto CANE. Ad ogni terminale è assegnata una caratteristica di *default*, che, cioè, si attiva in mancanza di ulteriori specifiche. Un'altra loro importante caratteristica è quella di poter dar luogo a rappresentazioni ricorsive, attraverso l'inserimento di un *frame* in un terminale di un altro *frame*. In tal modo è possibile procedere a rappresentazioni nidificate le une nelle altre, in modo da ottenere descrizioni gerarchiche (cioè, stratificate) e sempre più complesse delle situazioni da rappresentare, limitando il dispendio computazionale.

In termini generali, si può dire che un *frame* rappresenta un contesto e, dunque, va visto come una rappresentazione contestuale del concetto, passibile di un numero indefinito di specificazioni. Lo scopo per cui furono introdotti era quello di cogliere da una parte l'invarianza dei concetti e, dall'altra, la flessibilità cui vanno soggette le rappresentazioni concettuali di fronte alle varie istanze del concetto.

sia quanto più possibile *omogenea*, dove per omogeneità si intende la *possibilità di creare una corrispondenza strutturale fra le due rappresentazioni*.

Tre sono le considerazioni da fare in merito a questa impostazione. Innanzitutto, il programma non può operare se non è dotato di una conoscenza concettuale che gli permetta di costruire le descrizioni in modo che esse siano *sovrapponibili*. In altri termini, le figure nei riquadri devono essere descritte attraverso l'uso di una serie di concetti utili a rappresentare la figure, le parti delle figure e le relazioni fra le figure all'interno dei riquadri. Hofstadter propone di utilizzare una *rete semantica concettuale* e di procedere secondo la seguente euristica:

> [...] fare tentativi di descrizioni provvisorie per ciascun riquadro;
> metterle a confronto con le descrizioni provvisorie degli altri riquadri di ciascuna classe;
> ristrutturare le descrizioni:
> (i) aggiungendo informazione,
> (ii) eliminando informazione,
> (iii) vedendo la stessa informazione da un'altra angolazione;
> ripetere il procedimento finché non si trovi che cosa differenzia le due classi (*ivi*, p. 702).

Tale euristica procede sulla base di «regole esplicite» (*ibidem*) che indicano il modo in cui una *gerarchia di descrizioni*, da quelle più semplici a quelle più generali, viene composta. Naturalmente, a diversi livelli di descrizione corrispondono diversi concetti. Il livello base è quello dei *concetti primitivi* su cui edificare la struttura rappresentativa fino al livello dei concetti più astratti e delle «*descrizioni di descrizioni*, cioè *metadescrizioni*», che conducano all'individuazione di «un numero di caratteristiche comuni sufficiente a guidarci verso la formulazione di un profilo per le metadescrizioni» (*ivi*, p. 709). A questo livello le descrizioni diventano oggetto del programma stesso che cerca di equipararle sulla base di concetti più astratti. Altri due aspetti della componente euristica di questo modello teorico sono la "messa a fuoco" (*focusing*) e il "filtraggio" (*filtering*), le quali producono rispettivamente una descrizione «focalizzata su qualche parte del disegno del riquadro, escludendo ogni altra cosa» e una descrizione «che si concentri su qualche modo particolare di guardare al contenuto del riquadro e ignori deliberatamente tutti gli altri aspetti» (*ivi*, pp. 711-712). Il primo aspetto ha che fare con gli *oggetti percepiti* e il secondo con i *concetti interessati* (cioè, attivati) dall'operazione di costruzione della rappresentazione. Fra loro c'è una relazione di complementarietà.

Una seconda considerazione riguarda il fatto che tale programma si muove ancora nell'ambito del *simbolico*. Le rappresentazioni che costruisce della situazione, cioè degli insiemi di riquadri da porre in relazione *meta*-analogica attraverso *meta*-descrizioni, sono rappresentazioni simboliche che si avvalgono di una costruzione gerarchica *operata dal programma* di volta in volta nel corso dell'elaborazione e basata su *concetti primitivi* che vengono utilizzati nella fase di *pre-*

elaborazione. Hofstadter ne dà alcuni esempi (*ivi*, pp. 699-700), suddividendoli in quelli di primo livello – segmento, verticale, orizzontale, curva, nero, appuntito, piccolo, e così via – e quelli di secondo livello, che intervengono nella seconda fase pre-elaborativa – quadrato, cerchio, angolo retto, vertice, protuberanza, ecc. Come si vede, i primi si riferiscono a caratteristiche delle figure identificabili alla stregua di proprietà semplici, condivisibili da tutte le figure, i secondi sono già descrizioni di «*forme* elementari» che descrivono le figure stesse prese nella loro interezza o parti di esse dotate di una determinata forma. Il confine tra queste due categorie è, certamente, sfumato. Ciò che importa è che il passaggio dai primi ai secondi costituisce, in termini generali, il passaggio *dalle proprietà alle forme* per quanto abbozzate e grossolane queste siano.

Come si diceva, tali descrizioni sono effettuate attraverso il linguaggio della logica dei predicati e, quindi, in maniera fortemente simbolica. Le descrizioni sono *frame* i cui terminali corrispondono ai concetti primitivi di secondo livello e le metadescrizioni sono a loro volta *frame* che riportano, ad esempio, terminali relativi al tipo di concetti usati, ai concetti ricorrenti, ai nomi dei terminali delle descrizioni, ecc. In questo modo si ottiene quella *struttura concettuale astratta*, o anche lo *scheletro concettuale*, che gioca un ruolo essenziale nel mettere in correlazione i due insiemi di figure, sempre attraverso una messa in corrispondenza che si avvale della rappresentazione logico-predicativa, fino alla soluzione del problema meta-analogico di capire in che cosa differisce l'analogia fra i primi sei riquadri da quella dei secondi sei.

Un modello teorico di questo tipo ricorda molto da vicino, per il tipo di tecniche rappresentative impiegate, il programma ARCH di Winston che, sulla scia degli studi sulla visione compiuti in IA a partire dalla metà degli anni sessanta[3], progettò un sistema in grado di apprendere per generalizzazione induttiva a partire da esempi. Il programma di Winston (1975b) operava a partire da un serie di concetti primitivi per arrivare alla descrizione di un arco. Le proprietà e le relazioni attraverso cui il programma effettuava la descrizione erano pre-selezionate dal programmatore e la descrizione che produceva costituisce il tipico esempio di rappresentazione in forma simbolica, una lista di proprietà e relazioni, oggetto di attacco da parte dei primi critici dell'IA simbolica negli anni settanta. Il problema relativo alla conoscenza in dotazione a un programma è sorto, infatti, nel momento in cui la sua rappresentazione all'interno di un qualche programma di IA simbolica è stata considerata psicologicamente implausibile (da cui le numerose teorie anti-tradizionali sui concetti che sono state sviluppate negli ultimi trenta anni) e il programma accusato di non spiegare proprio ciò che la sua realizzazione avrebbe dovuto rendere chiaro. Tale critica era motivata dal fatto che, come afferma Dreyfus riferendosi ad ARCH, «l'attività di discriminazione, selezione, e dare un peso ad una limitata quantità di proprietà rilevanti è il risultato di esperienze ripetute nel tempo ed è il primo stadio dell'apprendimento. Ma poiché nel sistema di Winston il programmatore seleziona e

[3] Si vedano, tra gli altri, Guzman (1968), autore del programma SEE e Clowes (1971), Waltz (1972), continuatori su questo filone di ricerca dedicato alla visione artificiale.

soppesa i primitivi, il suo programma non ci dà alcuna idea su come un calcolatore potrebbe operare questa soluzione e assegnare quei pesi» (Dreyfus, 1981, p. 190).

In queste parole già si intravede la via che sarà presa di lì a poco dal connessionismo, che farà dell'apprendimento uno dei suoi cavalli di battaglia. Tuttavia, anche Hofstadter agisce per superare questo tipo di problemi e l'impasse che ne deriva. L'utilizzo di descrizioni basate su concetti e relazioni primitive è una caratteristica anche del suo modello teorico, la quale condurrebbe allo stesso *circolo vizioso esplicativo* del programma di Winston. Tale rischio viene evitato attraverso l'impiego di una rete semantica di concetti che, tuttavia, si differenzia da quelle tradizionali introdotte da Quillian (1968) gerarchicamente strutturate secondo un sistema classificatorio statico ad albero in cui ogni concetto è incluso in quelli di livello superiore e include quelli di livello inferiore. La rete di concetti proposta da Hofstadter è ancora una rete associativa, ma non rigidamente gerarchica. Ogni concetto è collegato a quelli con cui è in relazione attraverso legami predefiniti[4]. Hofstadter definisce il suo programma «eterarchico», perché «tutto ciò che è nella rete, cioè sia i nodi che gli archi», è importante; «non c'è niente nella rete che si trovi ad un livello superiore al resto». In altri termini è nell'elaborazione del programma, nella sua dinamica costruttiva delle descrizioni, che va cercata la componente gerarchica, via l'utilizzo di concetti primitivi *in base alle esigenze del programma* nel momento in cui svolge il proprio compito.

La dimensione eterarchica del modello teorico viene ampliata attraverso l'introduzione di una tecnica molto simile alla computazione asincrona e parallela degli attori di Hewitt[5]. La parte procedurale del programma, infatti, viene demandata ad una serie di agenti che, come gli attori proposti da Hewitt, possono interagire fra loro e «scambiarsi mutuamente messaggi complessi» (Hofstdater, 1979, p. 716)[6]. La computazione attraverso attori pone in atto forme di elaborazione competitiva e parallela. Da un punto di vista molto generale, si può dire che il programma viene scisso in sottoparti virtualmente indipendenti che possono procedere in maniera sincrona o asincrona scambiandosi informazioni relative al compito che stanno effettuando. La linearità dell'esecuzione dell'algoritmo si frammenta in tal modo in una serie di operazioni semi-indipendenti, nel senso che ogni attore agisce in base sia alle informazioni che possiede al momento presente, e che scambia dinamicamente con gli altri attori, sia alla particolare struttura di cui è costituito, lo specifico *software* che descrive le funzioni che è preposto a compiere.

Questa "eterarchia di procedure che si richiamano" sfrutta le potenzialità indefinitamente complesse dei messaggi che possono essere scambiati e si discosta in questo modo dall'operazione, usuale in informatica, della "chiamata di procedura". In questo modo, gli attori-agenti funzionano

[4] La cui variabilità e costruzione o distruzione costituisce uno dei punti più controversi, ma anche decisivi ai fini della simulazione dell'apprendimento.
[5] Si veda, ad esempio, Hewitt (1977).
[6] Tra essi, ad esempio, rientrano quello che Hofstadter chiama "Rico", ovvero riconoscitori di identità «continuamente in perlustrazione all'interno delle singole descrizioni e all'interno di descrizioni differenti, alla ricerca di descrittori o di altri elementi che si presentino identici più di una volta» (Hofstadter 1979, p. 702), al fine di operare ristrutturazioni della descrizione complessiva della situazione rappresentata.

alla stregua di «calcolatori autonomi, mentre i messaggi [che si scambiano] sono in qualche modo simili a programmi» (*ibidem*) che vengono interpretati dall'attore medesimo. Ciò che suggerisce Hofstadter è di potenziare a sua volta anche questo tipo di programmazione multiagente attraverso la fusione di unità procedurali e unità dichiarative di rappresentazione della conoscenza fino alla creazione di ideali macrounità di informazione e azione da lui chiamate *simboli* e risultanti dalla unione di «frame + attori» (*ibidem*). Sulla centralità e la complessità della nozione di simbolo in Hofstadter ritorneremo in seguito. Per ora basti dire che, nella proposta di modello teorico avanzata da Hofstadter, i simboli giocano il ruolo di *perni elaborativi* del programma, nel senso che costituiscono dei punti fissi attrattivi, non solo attorno ai quali ruota l'elaborazione, ma anche produttivi dell'elaborazione stessa. È molto importante sottolineare che la loro presenza nei modelli concreti che discenderanno da questa proposta è solo *virtuale e a un meta-livello rispetto a quello del programma*. In altri termini,il punto centrale è che la loro presenza non è esplicita nella sorgente del programma, bensì è frutto *emergente* dell'elaborazione.

Un'ultima considerazione in merito a questa proposta iniziale di modello riguarda l'obiettivo che intende conseguire. Ponendosi come punto di partenza quello della risoluzione dei problemi di Bongard, Hofstadter in realtà invita implicitamente a fare un passo oltre anche rispetto a quella che nel precedente capitolo abbiamo visto essere lo scopo dell'approccio subcognitivo, cioè la simulazione della capacità di percezione di alto livello. Infatti, tale tipo di problemi rientra in quello più generale di *riconoscimento delle forme* (*pattern*), fra le quali egli annovera a titolo di esempio anche «il riconoscimento delle facce [...], il riconoscimento di sentieri nei boschi e in montagna [...], la capacità di leggere senza esitazione testi composti in centinaia, se non migliaia, di caratteri tipografici differenti» (*ivi*, p. 719). Tali compiti rientrano all'interno del fenomeno della percezione in generale, non solo visiva, e quindi riguardano anche la percezione di basso livello. Alla simulazione di questo ultimo tipo di capacità è stata dedicata un'attenzione crescente proprio a partire dagli anni ottanta del secolo scorso, anche e soprattutto da parte dei nuovi approcci connessionisti alla simulazione delle capacità cognitive[7]. Dunque è nella spiegazione di come sia possibile l'integrazione fra alto e basso livello del fenomeno percettivo che va visto lo scopo finale dello sviluppo di modelli simili a quello appena descritto.

Sulle relazioni fra i modelli che abbiamo definito subcognitivi e il connessionismo ritorneremo in seguito. Per ora, è opportuno sottolineare che il fatto che Hofstadter porti in primo piano il problema della percezione di forme anche di basso livello, considerate alla base del meccanismo di descrizione e di metadescrizione il quale innesca il processo elaborativo che ha per oggetto rappresentazioni «*strutturalmente simili l'una all'altra*» (*ivi*, p. 702), apre la via e indica una direzione all'indagine di questi fenomeni con il considerarli *strettamente interconnessi* con i processi cognitivi di alto livello. Infatti, la capacità di operare descrizioni che *evolvono*

[7] Tuttavia, non va dimenticato il fondamentale contributo in questo campo da parte dell'approccio simbolico tradizionale all'IA dovuto a David Marr e di poco posteriore alla proposta hofstadteriana (Marr, 1982), che ha anche l'indubbio valore di aver costituito una pietra miliare nella metodologia delle discipline simualtive in generale.

dinamicamente su più livelli è costitutiva dell'esperienza percettiva di ognuno: «è molto probabile che le intuizioni ottenute vedendo e manipolando oggetti reali (pettini, treni, stringhe, blocchi, lettere, nastri adesivi, ecc.) svolgano un ruolo guida invisibile ma significativo nella soluzione di questi rompicapo» (*ivi*, pp. 714-15). Di conseguenza non stupisce che una delle principali assunzioni alla base del progetto hofstadteriano sia la seguente:

> [...] è sicuro che la comprensione di situazioni del mondo reale dipende fortemente dall'immaginazione visiva e dall'intuizione spaziale, cosicché disporre di un metodo potente e flessibile per rappresentare forme del tipo di quelle di Bongard può certamente contribuire all'*efficacia generale dei processi di pensiero* (*ibidem* [corsivo mio]).

La comprensione degli aspetti percettivi legati all'esperienza di eventi e situazioni spaziali (ma anche temporali; si pensi alla percezione musicale, basata su un ordinamento vincolato alla dimensione temporale), è imprescindibile, nella visione hoftadteriana, dalla comprensione dei processi cognitivi in generale, anzi ne costituisce uno degli aspetti basilari. Questo aspetto come vedremo ritornerà in tutti i modelli cognitivi basati su questa impostazione, costituendone uno dei minimi comuni denominatori teorici e mostrandone al tempo stesso le ampie implicazioni con una visione rappresentazionale esplicita, e, dunque, simbolica dell'IA e della simulazione dei processi di pensiero.

3.3 L'alfabeto come universo

3.3.1 Il progetto COPYCAT

Fra i modelli cognitivi che discendono dal modello proposto da Hofstadter quello che forse ha ricevuto più attenzione e più è stato discusso è COPYCAT (Mitchell, 1993; Mitchell, Hofstadter, 1994). Non è il primo ad essere stato sviluppato dal punto di vista cronologico, ma deriva da una serie di modelli sviluppati o giunti fino alla fase immediatamente precedente la realizzazione sul calcolatore, cioè grossomodo quella algoritmica, e progettati per operare su domini differenti. Da questi COPYCAT riprende alcune idee fondamentali relative alla sua componente algoritmica e computazionale e le trasporta nell'universo costituito dall'alfabeto.

Il problema prototipico che COPYCAT è in grado di affrontare è un problema di "risoluzione analogica" del tipo "se ***abc*** diventa ***abd***, che cosa diventa ***ijk***?", esprimibile anche, secondo la notazione comunemente usata, nel seguente modo:

abc => ***abd***, ***ijk*** => ?

L'espressione "risoluzione analogica" è, a ben vedere, fuorviante, perché, in realtà, non si tratta di una problema che ammette un'unica soluzione, ma diverse soluzioni più o meno plausibili. Tra esse, ad esempio, Mitchell riporta (Mitchell, 1993, p. 76): ***ijl***, ***ijd***, ***ijk***, ***hjk***, ***ijj***. Come si vede, si va da una risposta molto plausibile, in cui l'ultima lettera a destra viene trasformata nel suo successore, fino a risposte in cui nella stringa di lettere obiettivo (***ijk***) viene sostituito proprio lo stesso termine (***d***) della stringa di partenza trasformata; oppure viene ripetuta per intero la stringa obiettivo; o anche c'è un raddoppiamento della seconda lettera della stringa obiettivo (***j***); o, infine, viene cambiata la prima lettera della stringa obiettivo con quella che la precede (***h***).

Alcune di queste risposte possono apparire banali, altre insolite e quasi giocose, seppure il programma non sia stato progettato per esibire atteggiamenti umoristici. Altre, come ad esempio l'ultima, mostrano invece una certa sottigliezza, un certo grado di profondità concettuale nel costruire la risposta, in cui intervengono relazioni astratte come la *simmetria* e operazioni complesse come l'*inversione*. In realtà, il programma arriva molto spesso alla prima conclusione, quella più ovvia anche per essere umano, e soltanto in pochi casi alle altre. Tuttavia, il fatto che ci arrivi mostra una certa flessibilità di comportamento. Esiste, inoltre, la possibilità di influire sulle componenti strutturali del modello per far sì che il numero di certe soluzioni aumenti, anche se non in maniera considerevole. Per capire come, occorre considerare dapprima gli aspetti essenziali dell'architettura di COPYCAT.

In conformità alla TLCC esposta nel capitolo precedente, COPYCAT si compone di tre parti funzionali che corrispondono a tre componenti in grado di generale il *loop cognitivo* fra memoria a lungo termine, memoria a breve termine e conoscenza procedurale attiva (*ibidem*, pp. 31-73). Alla MLT corrisponde una rete detta "di Slittamento" (*Slipnet*), ovvero una rete semantica con alcune caratteristiche peculiari. Mentre nelle reti semantiche tradizionali *à la* Quillian, i nodi rappresentano i concetti e gli archi legami di inclusione o di appartenenza di classe, nella rete di slittamento sia i nodi che gli archi possono rappresentare concetti[8]. Ad essi corrisponde un certo grado di attivazione che varia conformemente alle fasi dell'elaborazione del programma. All'attivazione dei concetti nella rete si affianca un'altra proprietà, cioè la variabilità della lunghezza degli archi, che esprimono in questo modo la maggiore o minore vicinanza dei concetti che collegano.

La rete di slittamento è intesa in questo modo incorporare alcuni aspetti fondamentali delle capacità associative del pensiero. Infatti, la propagazione di attivazione dei concetti nella rete esprime di volta in volta la mutevole attenzione del programma nei confronti del compito che sta eseguendo e permette il passaggio, in maniera associativa, da un concetto all'altro, qualora se ne verifichino le condizioni. D'altra parte, la rete va pensata come un insieme di categorie prefissate definite al loro centro, ma sfumate quanto ai contorni della loro applicazione. Esse sono, cioè, *tipi* le cui istanze ne causano l'attivazione e che possono essere considerati anche come perni attorno a cui

[8] Se un arco è *etichettato*, esprime un concetto, la cui attivazione si riverbera sui nodi-concetti ad esso collegati.

si modifica la nuvola o alone di attivazione ad essi associata, costituito dai concetti prossimi. Questo meccanismo permette un forte potere rappresentazionale della situazione cui l'attenzione del modello viene rivolta, attraverso il meccanismo proiettivo rispecchiato nell'attivazione dei concetti coinvolti.

Altra componente fondamentale del modello è lo Spazio di Lavoro (*Workspace*) in cui avviene l'elaborazione a partire dal materiale immesso come input sotto forma di problema di analogia. All'interno dello spazio di lavoro il programma può compiere sei operazioni generali di costruzione di strutture (*ibidem*, p. 44): può *descrivere* i singoli oggetti presenti (lettere, ad esempio); creare *legami* (*bond*), cioè relazioni tra elementi; formare *gruppi* di elementi; istituire *corrispondenze* fra elementi di diverse stringhe del problema; produrre un *regola di trasformazione* che esplicita il cambiamento fra le prime due stringhe (***abc*** => ***abd***); fornire una *traduzione* della regola di trasformazione, che indica il modo in cui la stringa obiettivo dovrebbe cambiare. La traduzione è resa possibile proprio dalla condivisione di uno stesso nucleo concettuale astratto. Nell'interfaccia grafica sono soprattutto i legami ad esse visibili e rappresentati con archi fra i vari elementi della situazione percepita[9].

Se non è verosimile dire che tutta l'elaborazione del programma avviene in questa parte dell'architettura, si può altresì affermare che essa esplicita tutto il potere rappresentazionale del sistema, il quale risiede nell'illimitato potere di costruire rappresentazioni sulla scorta delle sei operazioni possibili, *applicabili ricorsivamente* alle strutture percepite sia ai singoli elementi già presenti nella fase iniziale, sia alle entità più complesse che scaturiscono nel corso dell'elaborazione. Lo spazio di lavoro, infatti, «è inteso corrispondere alla regione mentale in cui le rappresentazioni di situazioni sono costruite dinamicamente» (*ibidem*, p. 42). È già evidente, dunque, il tratto *mentalista-rappresentazionalista* che caratterizza questi modelli, e che, tuttavia, si accosta ad un'euristica architettonica ispirata alle dinamiche evolutive, come quelle presenti a più livelli nel dominio delle scienze biologiche. Ciò ha un riscontro nella particolare forma di simbolismo posta in essere dai modelli subcognitivi.

L'aspetto che lega questi sistemi alle dinamiche evolutive tipiche di molti fenomeni biologici è legato al forte parallelismo che ne caratterizza l'elaborazione, almeno nelle sue fasi iniziali. Questo, come si è visto, è reso possibile dall'utilizzo di microprocedure, cioè piccoli programmi semiautonomi, la cui attivazione, per "chiamata", determina l'andamento generale del sistema. Tali sotto-programmi, denominati codicelli (*codelet*) hanno un serie di funzioni specifiche differenziate, legate alle sei strutture costruibili nello spazio di lavoro. In termini molto generali, possono essere divisi in due tipi fondamentali: codicelli *bottom up* la cui "chiamata" è in qualche misura "spontanea", nel senso che operano a partire dal basso, indagando gli elementi dello spazio di lavoro senza alcuno scopo se non quello di costruire quante più strutture possibili sulla base della

[9] La loro dinamica continuamente in evoluzione è rappresentata attraverso il rafforzamento delle linee di collegamento e delle linee che circondano i gruppi degli elementi individuati dal programma.

loro funzione specifica (ad esempio, possono istituire un collegamento fra due lettere dello stesso tipo, o raggruppare lettere secondo un legame di successione); codicelli *top down*, immessi nell'elaborazione a seguito dell'attivazione dei nodi della rete concettuale e, quindi, in una certa misura *vincolati* allo spazio concettuale nella loro costruzione di strutture (ad esempio, sono in grado di creare un gruppo di una certa lunghezza se il concetto relativo a quella lunghezza è attivo nella rete di slittamento). Tuttavia, questo modo di descrivere la parte attiva del programma non deve trarre in inganno sul modo in cui effettivamente vengono scelti i codicelli. La loro attivazione è regolata, infatti, in base ad una variabile che esprime la loro *urgenza* e il cui valore dipende, ad eccezione di quelli *bottom up* presenti nella fase iniziale dell'elaborazione, sia dall'attivazione di codicelli precedenti, che esprimono dunque una valutazione sul tempo maggiore o minore che dovrà intercorrere prima del successivo utilizzo di un medesimo codicello, sia dall'attivazione dei nodi nella rete. Ad una maggiore attivazione nella rete corrisponde, infatti, una maggiore urgenza di attivazione del codicello, così che l'andamento generale dell'elaborazione del programma sia progressivamente sempre più condizionato dalle pressioni concettuali attive a mano a mano che il programma percepisce gli elementi della situazione esaminata.

A un livello di dettaglio ancora maggiore le microprocedure possono essere distinte in tre tipi applicabili alla costruzione di ogni struttura (descrizione di un oggetto, formazione di gruppi, corrispondenze e regole, ecc.): esploratori, valutatori, costruttori. Tale distinzione indica anche che ogni struttura deve passare attraverso questi tre stadi prima di essere costruita e non è detto che ogni volta che il primo o il secondo stadio siano raggiunti, automaticamente il terzo sia prodotto. Perciò, nel problema che abbiamo preso precedentemente in considerazione (***abc*** => ***abd***, ***ijk*** => ?), la creazione di un collegamento tra la ***a*** della stringa iniziale e la ***a*** della stringa modificata deve sottostare alla seguente «catena tri-microprocedurale:

- un codicello esploratore sceglie probabilisticamente un oggetto o alcuni oggetti su cui costruire la struttura, e si chiede "C'è una qualche ragione per costruire questo tipo di struttura?"
- Se la risposta è sì, un codicello valutatore-di-solidità si chiede "La struttura proposta è abbastanza forte?"
- Se la risposta è sì, un codicello costruttore prova a costruire la struttura, lottando contro i competitori se necessario» (*ivi*, p. 64).

Come si vede, la struttura tripartita di ogni procedimento di costruzione ne garantisce il vaglio entro i limiti delle risorse computazionali del sistema. Il procedere parallelo dell'elaborazione fa sì che non c'è un univoco «sentiero elaborativo», ma esso è piuttosto il risultato complessivo di «un insieme di passi che conducono a una risposta, a cui partecipa un ampio numero di codicelli e di strutture» (*ivi*, p. 65).

La combinazione di funzioni *top down* e *bottom up* ed esplorativo-valutativo-costruttive di differenti strutture fa sì che le microprocedure del programma siano dell'ordine di una ventina, alcune aventi la forma di funzioni mono-argomentali che possono essere funzioni matematiche in senso proprio e assegnare un valore numerico (ad esempio, le microprocedure valutative, che esprimono con un valore numerico la valutazione effettuata); altre che si caratterizzano come "funzioni senza argomento" nel senso che il loro scopo è quello di ritrovare elementi nello spazio di lavoro e di porli in corrispondenza in base alla loro funzione specifica (ad esempio, quella di costruire legami). In un certo senso qui si evidenza un'ambivalenza nella nozione di *funzione*, da una parte intesa come concetto matematico che fa corrispondere valori ad oggetti presi come argomenti, ovvero mette in corrispondenza elementi di differenti insiemi; dall'altra in senso operativo, come attuazione di collegamenti nello spazio degli elementi percepiti. Questa ambivalenza, basata sull'analogia fra i due tipi di funzione in quanto *operazioni di messa in corrispondenza*, è uno dei principali elementi a favore di una considerazione semantica, e non solo sintattica, dell'attività del sistema. Il programma, infatti, sviluppa una sorta di *comprensione* della situazione che è espressione sia della valutazione che fa degli elementi e delle strutture costruite[10], sia dell'attivazione dei concetti della rete di slittamento, la quale influisce direttamente sulle operazioni compiute dal programma.

Tuttavia, l'assegnazione di un valore alle strutture sembra far propendere verso una mancanza di plausibilità psicologica del modello, come viene segnalato anche dall'autrice del programma:

> Il ruolo delle funzioni per calcolare le forze delle strutture è dunque non di proporre meccanismi psicologici dettagliati del modo in cui i valori di forza sono computati, bensì piuttosto di produrre numeri plausibili che possano essere utilizzati nei meccanismi che *stiamo* proponendo, così come nell'esplicitare le pressioni che sono coinvolte nell'insorgere di tali numeri. (*ivi*, p. 62)

Interpretando il passo e generalizzandolo anche ad altri aspetti dell'elaborazione, può esserne tratta un'assunzione di fondo che taglia trasversalmente la reale simulatività di questi modelli, ovvero che la valenza effettiva delle variabili numeriche è quella di essere tasselli dell'elaborazione utili dal punto di vista computazionale a tradurre in termini implementativi i concetti (ovvero, le pressioni operative da essi esercitate) *realmente impiegati*, e dunque posseduti, dal programma.

[10] «Ogni struttura ha un *forza* che varia nel tempo e che misura la sua qualità, ed ogni oggetto ha una *felicità* che varia nel tempo e che misura il grado di bontà del suo adattamento all'interno dell'insieme corrente di tutte le strutture» (Mitchell, 1993, p. 58). La composizione dei valori di forza e infelicità (l'inverso del valore di felicità) esprime un altro tipo di dato numerico che ha un notevole peso nella elaborazione: il valore di *salienza* che ogni struttura ha dal punto di vista del programma. Si può dire che tanto più il programma percepisce un elemento come stabile e felice, tanta meno attenzione gli rivolge. Naturalmente, tale funzione è ancora un esempio di emergenza nel corso dell'elaborazione e la *relazione causale*, in questo come in tutti gli altri casi, va vista in primo luogo nella direzione dati-programma e non viceversa. In altri termini, non è il programma a decidere in via preliminare su quali elementi concentrarsi, ma il fatto che determinati elementi abbiano una certa salienza sta a significare che, almeno in buona parte, il *focus* attentivo del sistema si è già concentrato su quei dati nella fasi iniziali in cui la quantità di *casualità* dell'elaborazione è massima.

L'elaborazione di COPYCAT procede, in accordo con i principi esposti nel capitolo precedente, in modo parallelo, mettendo in atto un processo che va dallo stocastico al deterministico. Ad una iniziale fase esplorativa in cui tutti gli elementi vengono presi in considerazione e "testati", segue la creazione di strutture percettive sempre più stabili e sempre più coerenti tra loro. Tanto più gli elementi vengono collegati fra loro in un'unica costruzione cui corrisponde una struttura concettuale complessiva, tanto minore è l'andamento casuale dell'elaborazione e le microprocedure attivate saranno tutte concentrate sulle strutture più grandi e stabili al fine di ottenere un disegno unitario, tradotto poi in una *regola di trasformazione*. L'elaborazione parallela e non deterministica del programma è garantita dal meccanismo di selezione dei codicelli, il valore di urgenza dei quali stabilisce la probabilità della loro "chiamata" ed è funzione sia di pressioni *bottom up* che *top down*, cioè dell'esecuzione di precedenti microprocedure nello Spazio di Lavoro e dell'attivazione della rete concettuale, le quali entrambe assegnano il valore di urgenza dei nuovi codicelli. Appare immediatamente evidente che in questo modo viene attuata una *selezione di natura emergente* nell'insieme dei percorsi di elaborazione del sistema fino all'esito completamente deterministico di un solo percorso elaborativo, corrispondente a un unico punto di vista espresso dalla regola. COPYCAT procede all'attuazione di questo meccanismo attraverso un ciclo applicato *ricorsivamente* agli elementi dello spazio percettivo ogni volta che una microprocedura viene attivata[11].

Facciamo alcuni esempi. Ritornando al problema iniziale (***abc*** => ***abd***, ***ijk*** => ?), ciò che il programma farà, diretta conseguenza del modo in cui "vede" la situazione che sta analizzando, sarà quello di creare un collegamento fra le due ***a***, rispettivamente, della stringa di partenza e di quella modificata; e ancora fra le due ***b***. A quel punto collegherà ***c*** e ***d***, e si appresterà a creare ponti fra la stringa di partenza e quella obiettivo, ad esempio fra ***a*** ed ***i*** per il fatto che occupano la stessa posizione nella stringa, e così via. L'effettuare i collegamenti fra la stringa iniziale e quella modificata lo porterà a esprimere in una regola (in forma di proposizione in linguaggio naturale) il modo in cui la prima stringa cambia nella seconda. La regola sarà, dunque, qualcosa del tipo:

i) *Rimpiazza la categoria di lettera della lettera più a destra con la sua successiva*.

A partire da questo punto, il programma cercherà di applicare la regola di cambiamento alla stringa obiettivo per ottenere la sua trasformazione, che deve essere analoga a quella delle due stringhe,

[11] Il ciclo generale dell'elaborazione, che incorpora quello delle singole microprocedure, è il seguente:
«Fino a che una regola non è stata costruita e tradotta, ripeti:
Scegli un codicello e rimuovilo dalla Scatola dei codicelli.
Esegui il codicello scelto.
Se *N* codicelli sono stati eseguiti, allora:
aggiorna la Rete di slittamento;
imposta codicelli *bottom up*;
imposta codicelli *top down*.
Infine, costruisci la risposta in accordo alla regola trasformata» (Mitchell, 1993, p. 72).

iniziale e modificata[12]. In tal modo, anche il processo che porta alla strutturazione della regola di modificazione incorre in un processo di rimandi fra ipotesi di regole e possibili adattamenti al rapporto individuato fra la stringa obiettivo e quella ipotetica finale, cioè ancora una volta fra concetti e strutture percettive.

Altri esempi significativi sono dati da Mitchell (1993, pp. 75-169) e di alcuni di essi sono proposte varianti, utili a testare il potere del programma su semplici variazioni delle situazioni in oggetto. Non potendo ripercorrerli tutti, ne citiamo ancora due, per chiarire ulteriormente più che le effettive potenzialità del programma, quali sono gli obiettivi che gli autori si pongono con la sua realizzazione.

Il primo è un problema del tipo:

abc => ***abd***, ***mnnooo*** => ?

In questo caso il programma dovrebbe essere in grado di fornire due risposte sufficientemente plausibili e dotate di una certa profondità concettuale, oltre a quelle più "superficiali". La prima è ***mnnppp*** che mostrerebbe il fatto di aver percepito non solo l'ultima lettera della stringa obiettivo, bensì tutto il gruppo formato da istanze dello stesso tipo di lettere, come la parte che deve essere trasformata (nel caso in questione con il successivo tipo, o la successiva categoria, di lettera). La seconda risposta dovrebbe essere ***mnnoooo*** che, attuando un cambiamento nella lunghezza del gruppo, denoterebbe una messa in corrispondenza, attraverso il concetto di successione, dell'insieme ordinato delle lettere con quello dei numeri naturali[13]. Appare chiaro che questo secondo tipo di risposta è dotato di una grado maggiore di profondità concettuale, e, di conseguenza, che il sistema ha operato una percezione *più astratta* della situazione, ponendo in corrispondenza due tipi di relazione sulla base di una più generale meta-relazione d'ordine, espressa dalla coppia di concetti dicotomici "*successore*/*predecessore*". COPYCAT, di fatto, è in grado di dare entrambe queste risposte[14] e la maggiore astrattezza dell'una rispetto all'altra è testimoniata dalla frequenza molto minore con cui il programma trova la seconda risposta rispetto alla prima. Nel fare questo, il sistema esibisce una capacità psicologicamente molto plausibile dal punto di vista umano, relativa alla maggiore difficoltà nell'utilizzare concetti più astratti e nel percepire il problema di analogia che sta affrontando a un livello più profondo, *meta-concettuale*, ovvero come

[12] Si ricordi ancora una volta che il processo è soltanto a posteriori descrivibile attraverso un ordine temporale determinato. Il parallelismo delle microprocedure fa sì che tutti i processi di collegamento, compresi quelli di raggruppamento e quelli di produzione di regole, non avvengano secondo un ordine stabilito, ma secondo il meccanismo probabilistico descritto in precedenza.

[13] COPYCAT non è fornito di un generatore della successione infinita dei numeri naturali. La sua rete semantica comprende solo l'insieme dei primi cinque numeri, ma interconnessi secondo una relazione d'ordine che rispecchia la loro successione nella serie dei numeri naturali: appunto 1, 2, 3, 4, 5.

[14] Si veda l'esempio riportato in Mitchell (1993, p. 163), nella sezione dedicata alle *variazioni sul tema* rispetto ad alcune categorie di problemi.

problema su concetti e non soltanto su oggetti o categorie di oggetti definite in maniera puramente estensionale per relazione di appartenenza fra istanza e tipo.

Dal punto di vista dei processi conoscitivi simulati l'aspetto più interessante dell'architettura del programma è, probabilmente, la rete concettuale e il modo in cui il sistema *ha* conoscenza. L'espressione di regole di trasformazione in linguaggio naturale non denota una capacità linguistica *human-like* da parte del programma. La produzione di regole di trasformazione avviene attraverso il riempimento di sagome (*template*) preformate. D'altra parte, non rientra fra gli obiettivi di COPYCAT quello di esibire capacità di comprensione e generazione (sintattica) di espressioni del linguaggio naturale. Piuttosto, è il modo in cui gli spazi vuoti delle sagome vengono *appropriatamente* riempiti dal programma a costituire uno degli scopi rilevanti dell'intero approccio subcognitivo. In questo, un ruolo fondamentale è giocato dalla rete semantica, che, come nel modello teorico formulato da Hofstadter per la risoluzione dei Problemi di Bongard, assume connotati eterarchici. Per tale ragione, occorre considerare in maniera più specifica gli aspetti principali della sua struttura.

La rete semantica di COPYCAT contiene una cinquantina di concetti, tra cui i principali sono i 26 tipi di lettere, i numeri da 1 a 5, relazioni di posizione ("*left*", "*right*"), concetti che esprimono tipi di legami ("*predecessor*", "*successor*", "*sameness*"), tipi di gruppi, categorie ("*letter*", "*group*") e una serie di concetti che corrispondono a meta-categorie descrittive (ad esempio, "*letter-category*", "*string-position*", "*object-category*", "*alphabetic-position*", "*bond-category*") e due tipi di meta-nodi, "*identity*" e "*opposite*", per esprimere relazioni fra i nodi stessi all'interno della rete: essi, cioè «etichettano relazioni nella Rete di Slittamento (Mitchell, p. 48). I concetti "base", cioè meno astratti, sono abbastanza intuitivamente quelli che rappresentano i tipi di lettere, l'idea "platonica" che si attiva nella rete in corrispondenza della lettera percepita nello Spazio di Lavoro. Tali concetti sono, in realtà, più semplici dei concetti di lettera che un essere umano possiede. Essi sono definiti solo secondo il *matching* che possono avere con una precisa istanza e dalle relazioni con le altre lettere (prossime) nell'alfabeto. Nei nodi non è, ad esempio, racchiusa la ricchezza delle varie forme in cui si può dare una stessa categoria di lettera, e neppure tutte le relazioni esistenti fra una lettera e tutte le altre, ma solo quelle che esprimono il suo rapporto con quelle vicine, quella precedente e quella successiva.

La rete è dotata di una serie di accorgimenti interessanti. A fronte del fatto che i nodi hanno un differente grado di semanticità, «un numero che rappresenta la generalità o l'astrattezza dei concetti implicati» (Hofstadter, Mitchell, 1988, p. 97) il quale esprime un vincolo invariabile prefissato dal programmatore, il modulo concettuale presenta una serie di proprietà dinamiche, cioè varianti nel tempo. Ciò accade, come si è visto, ad opera delle microprocedure che rendono la rete *modello della situazione percepita* attivando i concetti della rete. Tale attivazione agisce sulla rete in due modi: per propagazione ai nodi vicini e modificando gli archi. La propagazione ai nodi vicini, cioè collegati, simula la capacità umana di possedere un concetto non come un'unità a sé stante, ma

come insieme delle sue *proprietà descrittive* (che tipo di concetto è) e *contestuali* (a quali concetti dello stesso tipo si lega). Entrambi i tipi di proprietà sono espressi da collegamenti della rete, rappresentati da meta-nodi, i quali possono essere di diverso tipo: *di categoria, di istanza, di proprietà, di slittamento* e *laterali*. In particolare sono interessanti questi ultimi due tipi.

Un tipo di meta-nodo che favorisce lo slittamento è "*opposite*", il quale mette in relazione concetti oppositivi. La rete non solo è in grado di *propagare* da un nodo all'altro l'attivazione al fine di costruire una regola sufficientemente o riccamente esplicativa, ma l'attivazione del meta-nodo è in grado di modificare la lunghezza del collegamento, la quale diminuisce all'aumentare del valore di attivazione del meta-nodo, favorendo lo *slittamento* tra due nodi collegati[15]. Un collegamento di tipo laterale, invece, esprime una «relazione semantica non gerarchica» (Mitchell, 1993, p. 50). Esempi di meta-nodi di questo tipo sono "*predecessor*" e "*successor*", che instaurano relazioni d'ordine *orizzontale*, cioè intra-livello, e collegamenti fra nodi che esprimono direzioni e posizioni spaziali ("*left*", "*leftmost*"), e che, dunque, appartengono a una medesima categoria sovra-ordinata.

Infine, va osservato come la struttura gerarchica multi-livello si rispecchia anche in una differenza sottile ma determinante nella rappresentazione del concetto di uguaglianza nella dotazione epistemica del programma. Mentre "*sameness*", infatti, esprime una relazione tra oggetti dello spazio percettivo ed è, dunque, un nodo fra gli altri, "*identity*" è un meta-nodo che sta ad indicare una relazione fra nodi nella rete. La gerarchia presente all'interno della rete rispecchia quella del grado di astrazione di cui sono dotati i concetti, che gli autori del programma identificano con la *semanticità*. Ad un grado di semanticità maggiore è connesso un processo di decadimento più lento, in conformità all'idea secondo cui una volta giunti, con un dispendio computazionale più elevato in termini cognitivi, a un livello di analisi più astratto, se ne è influenzati più in profondità e più a lungo rispetto all'influenza di ciò che possiamo percepire di più superficiale nella situazione concreta come divergente rispetto a un'interpretazione che faccia uso di tali concetti. Aspetto fondamentale di questa architettura concettuale è che il passare da concetti meno astratti a concetti più astratti corrisponde al duplice passaggio da una maggiore a una minore dipendenza dal dominio, e da una minore a una maggiore dipendenza dal contesto. Sembra, perciò, non ingiustificato affermare che il tipo di rappresentazione della conoscenza che la rete riesce a simulare va verso una maggiore semanticità con l'aumentare dell'astrazione, perché allontanandosi dagli elementi percepiti come istanze categoriali si arriva a un insieme di concetti che hanno bisogno di una forte attivazione contestuale (in riferimento al contesto concettuale) per essere a loro volta attivati. Questo tipo di relazione di graduale complementarietà potrebbe essere raffigurata attraverso lo schema della figura 3.2.

[15] Lo slittamento è un passaggio massiccio di attivazione che, per così dire, si riversa da un nodo all'altro grazie all'elevato valore di attivazione del meta-nodo che *costituisce* e, di conseguenza, *etichetta* il loro collegamento. Ad esempio, è quello che può accadere fra i nodi *first* e *last*.

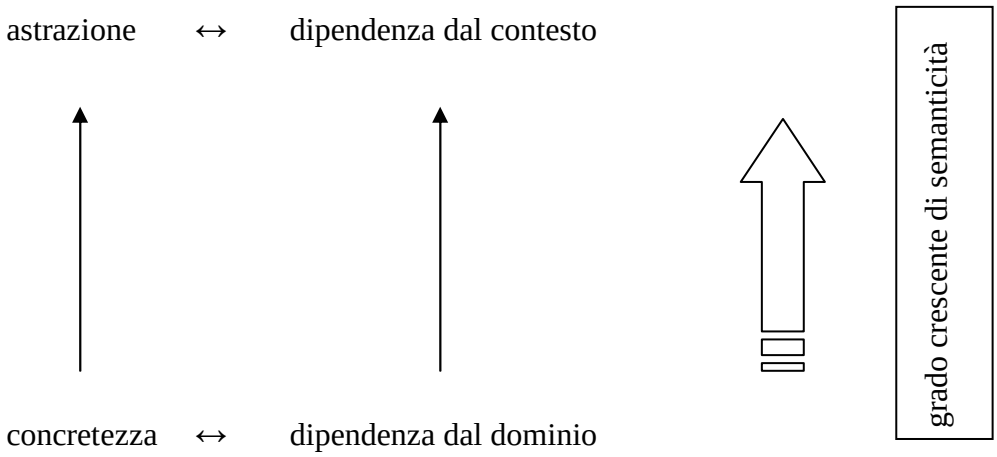

Fig. 3.2

Tale schema rappresenta l'andamento generale dell'elaborazione del modello dal punto di vista epistemico o conoscitivo. I concetti sono soggetti ad attivazione e il *pattern* generale di attivazione dei concetti esprime la *conoscenza*, *dinamicamente intesa*, che il modello ha della situazione che percepisce. Non esiste un'unità di controllo centrale che determina quale direzione intraprendere di volta in volta sulla base della rappresentazione creata e posseduta da COPYCAT. Piuttosto, è la rete che, fin dalle fasi immediatamente seguenti quelle iniziale, comincia a determinare l'andamento dell'elaborazione, come, per usare una metafora, una serie di filtri sempre più complessi che *dall'alto* stabiliscono ciò che deve essere percepito e ciò che deve essere tralasciato. Mitchell invita a considerare questo processo come qualcosa di analogo a ciò che avviene nelle *cell assembly* hebbiane[16]: «in un senso molto approssimativo, un nodo della Rete di Slittamento può essere pensato come analogo ad una *cell assembly* [...]. Il livello di attivazione del nodo corrisponde alla percentuale di neuroni attivi nella *cell assembly*. Se abbastanza neuroni sono attivi in una *cell assembly*, l'intera *assembly* tenderà a diventare attiva come risultanza delle connessioni ravvicinate fra le cellule» (Mitchell, 1993, p. 48). Rimane, tuttavia, una cospicua differenza relativa al fatto che i nodi della rete non rappresentano neuroni, bensì concetti. La corrispondenza è, perciò, da vedere tra *cell assembly* intere e singoli nodi della Rete di Slittamento. Il superamento di una certa soglia del valore di attivazione di un nodo ne causa la piena attivazione[17]. Questo ultimo meccanismo, a prima vista banale, è in grado di modellare la funzione psicologica dell'attenzione cosciente, caratterizzata da una forte *discontinuità* fra i momenti della sua presenza e della sua assenza.

In definitiva, si può affermare che la rete, determinando l'elaborazione del sistema, costituisce la *black box* rappresentazionale dinamica del programma e che è su di essa che si fa l'assunzione

[16] Si ricordi che per Hebb (1949) le assemblee cellulari sono, in termini molto generali, *cluster* di neuroni che si attivano oltre un certo valore di soglia risultante dalla somma dei valori di attivazione dei singoli neuroni da cui sono costituite.
[17] La scala dei valori di attivazione è in genere compresa fra 0 e 100. Il valore di soglia è quello intermedio, 50, superato il quale il nodo viene portato a 100.

funzionale più forte. Infatti, dal modo in cui essa è costruita e dai vincoli cui è sottoposta la sua evoluzione dinamica, dipende il comportamento generale di COPYCAT. L'assunzione, o postulato, che qui è in gioco è relativa alla funzione rappresentazionale, negata da approcci in modo diverso eliminativisti, come, ad esempio, quelli connessionisti. Sono i concetti e la loro interrelazione a determinare il comportamento non solo del sistema, ma della sua capacità percettiva (di alto livello). Quest'ultimo aspetto viene considerato, secondo tale impostazione, una forte prova a sostegno della esistenza dei concetti e della loro precipua *funzione rappresentazionale*.

Un ultimo elemento determinante dell'architettura di COPYCAT è la già menzionata variabile *temperatura*, cui è relegata la funzione di esprimere il valore di soddisfazione del programma riguardo alle strutture percettive che costruisce progressivamente, alle soluzioni date, e alle regole di trasformazione della stringa obiettivo nella stringa risposta. La temperatura è, dunque, strumento valutativo dell'elaborazione *dall'interno per l'esterno*, nel senso che è il modello stesso ad avere incorporata una funzione numerica, collegata all'attività delle microprocedure che presiedono alla costruzione delle strutture, i cui risultati concorrono a determinare dinamicamente il valore della temperatura[18]. In tale funzione di autoregolazione, corrispondente ad una sorta di autovalutazione, va vista la principale forma di auto-osservazione (*self-watching*) sviluppata in questo modello.

La variabile temperatura, tuttavia, non esaurisce in questo le sue funzioni. Essa è determinante nel contribuire a evitare il blocco del sistema di fronte a quelle tipologie di problemi che comportano un ostacolo (il così detto "*snag problem*"). Il secondo esempio utile a capire il funzionamento e le potenzialità di COPYCAT, riguarda questo tipo di problemi. Si consideri il seguente quesito di analogia:

abc => ***abd***; ***xyz*** => ?

Il programma, nell'analizzare la seguente situazione per darne una formulazione in termini di strutture percepite si troverà di fronte, progressivamente, a gruppi di successori, ponti fra lettere uguali e ponti fra lettere nella stessa posizione. Tuttavia, poiché non esiste il successore di ***z***[19]**,** ad un certo punto il programma si arenerà poiché, avendo composto una serie di strutture e non riuscendo a percepire altro, non è in grado di trovare la soluzione, pure con un abbassamento consistente della temperatura dovuto al formarsi di strutture stabili. Infatti, la variabile temperatura avrà un valore tanto più basso quanto più la stabilità delle strutture si riverbererà in un grado maggiore di "felicità"

[18] La funzione che calcola la temperatura è una funzione che prevede la soddisfazione di due variabili. Una sua forma possibile è $T = (0.8 * k) + (0.2 * p)$ dove k è la media del valore dell'infelicità di tutti gli oggetti, pesata in base alla loro importanza, e p è uguale a 100 meno il valore di forza (o stabilità) attribuito alla regola di trasformazione. Il valore della temperatura serve a determinare il valore della funzione mono-argomentale, che esprime l'urgenza di chiamata della singole microprocedure (Mitchell, 1993, p. 254). Poiché anche il grado di casualità che guida l'elaborazione e quello di decadimento delle microprocedure dipendono dal valore della temperatura, risulta evidente il ruolo centrale rivestito da questa variabile.

[19] Nella rete dei concetti non è presente l'idea di una circolarità che connetta la prima e l'ultima lettera dell'alfabeto.

per strutture già trovare che, in qualche modo, sono viste "sistemare" o "ordinare" la situazione. Ciò che la temperatura fa a questo punto è aumentare di nuovo generando in questa maniera probabilismo, attraverso il livellamento delle urgenze dei codicelli, fra i quali saranno presenti anche quelli distruttori. Ciò favorisce, ma solo in alcuni casi, il ritrovamento della seguente risposta: ***wyz***, che mostra come il rapporto fra le prime due stringhe venga percepito e riadattato in maniera simmetrica alle seconde due. Il fatto che la risposta sia rara è dovuto, come al solito, al maggiore grado di astrazione che caratterizza i concetti coinvolti, un livello cui non sempre il sistema arriva.

Anche i solutori di problemi umani mostrano di avere la capacità di azzerare tutto di fronte a un vicolo cieco e ricominciare da capo, abilità assente in alcuni casi nel mondo animale. Hofstadter definisce quest'ultima mancanza "*sphexiness*" (Hofstadter, 1985b, p. 529) dal noto esempio della vespa *Sphex*. Essa trascina il cibo davanti al nido, entra a vedere se tutto è a posto, esce e lo porta dentro. Se il cibo viene trascinato via, la vespa lo riporta all'entrata, non dentro, va ancora a vedere com'è la situazione e poi lo trascina dentro. Ad ogni spostamento del cibo segue un'identica procedura, all'infinito, o, meglio, fino all'esaurimento delle risorse corporee. L'uscita dall'atteggiamento di *sphexiness* costituisce un *salto di livello*, nel senso che la situazione viene considerata da un punto di vista superiore rispetto a quello corrente, al fine di evitare l'*impasse* di un atteggiamento circolare, che ha molte analogie con il *loop* informatico.

C'è, peraltro, una sorta di differenza fra superamento dell'ostacolo e uscita dal *loop*, che porta a considerare due modi diversi in cui il sistema può andare in *loop*. Il primo, che si può considerare minore, è la ripetizione di una stessa sottoparte del cammino di soluzione fino al medesimo punto più di una volta, e al limite tendente a infinito (in un caso concreto di una macchina dotata di energia e dispositivi di memoria finiti, fino all'esaurimento delle risorse; nel caso di una macchina o di un sistema ideale dotato di risorse infinite, senza una terminazione). Questo è un problema tipico in informatica, che riguarda la differenza fra un algoritmo corretto e uno non corretto, da non confondere con il Problema della Fermata di Turing, relativo alle possibili soluzioni di classi di funzioni. Questo è anche il caso in cui si trova COPYCAT nel momento in cui non riesce ad aggirare l'ostacolo, continuando a esplorare le strutture formate attraverso l'applicazione, ripetuta a intervalli regolari, degli stessi codicelli esplorativi. La funzione della variabile temperatura è proprio quella di forzare il programma a rompere le strutture e dunque a cambiare radicalmente la sua visione del mondo corrente, con la speranza di arrivare a descrivere la situazione in modo totalmente diverso e proficuo. Va considerato, peraltro, che il chiudersi in un vicolo cieco da parte del sistema deriva proprio dalla mutua non interferenza, ottenuta tramite tecniche di pianificazione gerarchica, delle microprocedure in azione, pena il rischio di cadere nell'impossibilità di creare strutture sufficientemente stabili per il ritrovamento di una soluzione. In altri termini, ciò che mostra il sistema è una sorta di rapporto di proporzionalità inversa fra determinismo e creatività, spingendo a considerare quest'ultima come largamente derivata dalla casualità. In questo modo si esprime Hofstadter al riguardo, in un saggio dedicato alla creatività di "chi ripete a pappagallo" (*copycat*):

> A molte persone è estremamente poco congeniale l'affermazione secondo cui un'intelligenza maggiore può scaturire dal prendere decisioni *casuali* (*random*) piuttosto che dal prenderne di *sistematiche*. Infatti, quando l'architettura di COPYCAT è descritta in questo modo, appare priva di senso. Non è forse sempre più saggio scegliere l'azione *migliore* piuttosto che scegliere *a caso* (*at random*)? Tuttavia, come in numerose discussioni sulle menti e i loro meccanismi, questa apparenza di insensatezza è un'illusione causata da una confusione di livelli. (Hofstadter, 1994, p. 420)

La presa di posizione a favore di una casualità come "motore della creatività" abbraccia tutta l'impostazione dell'approccio subcognitivo e in qualche maniera pone un punto di vista alternativo sia agli approcci dell'IA tradizionale fondati sulla razionalità perfetta (von Neumann, Morgenstern, 1944), sia a quelli basati sulla razionalità limitata (*bounded rationality*) proposti da Simon per l'analisi delle teorie del comportamento economico (1955, 1987), che, come è noto, a partire dalla fondazione dell'IA come disciplina autonoma ne hanno costituito l'assunto di fondo della tradizione psicologista. L'idea alla base dei modelli subcognitivi è che un certo grado di produttività è legato inscindibilmente ad operazioni casuali compiute all'interno del sistema cognitivo. Ciò costituisce un elemento di differenziazione e di svolta rispetto ai modelli precedenti da parte di quelli di derivazione hofstadteriana. Il caso, come le scienze evolutive ipotizzano trovando numerose conferme alle loro teorie, è produttore di creatività in quanto, per definizione, produttore di novità. Naturalmente senza i vincoli imposti dai processi di cristallizzazione (che corrispondono, da un punto di vista matematico, a punti di attrazione in un sistema dinamico) non si avrebbe la fissazione della novità immessa nel sistema. La funzione della temperatura è proprio quella di regolare la *quantità di cristallizzazione* del processo, che, lasciato alle sue normali conseguenze, è per definizione *incline alla costruzione* (si ricordino le funzioni specifiche delle microprocedure) e solo nel caso di un blocco (*loop* di primo tipo) può decidere di re-immettere causalità, cioè di livellare le urgenze probabilistiche delle microprocedure, ovvero delle *azioni possibili*, e di inviare particolari tipi di codicelli con il compito di distruggere le strutture già formate.

Esiste, tuttavia, un secondo tipo di *loop* in cui può cadere il sistema, che è di portata più ampia e relativo al caso in cui il programma compia un numero indefinito di volte lo stesso intero percorso di costruzione di strutture a partire dalle medesime fasi iniziali dell'elaborazione, anche se è in grado di uscire fuori dal *loop* definito in precedenza, che abbiamo definito "minore" perché circoscritto all'analisi delle stesse strutture già costruite. COPYCAT, fa notare Mitchell, non è attrezzato per affrontare questa situazione, che, come nel caso della vespa *Sphex*, richiede la capacità di memorizzare la sequenza delle azioni compiute e di poter disporre di questo tipo di informazioni ad ogni nuova elaborazione del problema. In altri termini, non è detto che l'aumentare della temperatura e l'immissione di causalità porti all'aggiramento dell'ostacolo. Ciò si potrebbe verificare solo in un numero molto ristretto di casi e il fatto di operare una decostruzione delle

strutture non impedisce al programma di ricreare le stesse strutture nello stesso modo, bensì gli permette di uscire dal *loop* di primo tipo più circoscritto che consiste nell'analisi ripetuta e infruttuosa delle strutture costruite.

Il limite cui va soggetto il sistema nel secondo caso è strettamente connesso con la mancanza di capacità di *learning* sulle sue azioni. COPYCAT, infatti, non incorpora nella sua architettura parti dedicate a questo scopo, ad esempio sfruttando le tecniche tipiche dell'apprendimento automatico. Al più, COPYCAT, può essere considerato un modello che apprende in relazione al suo potere adattivo rispetto alla situazione percepita, se si considera come apprendimento il modo in cui la rete *modella* o *rappresenta* la situazione stessa, il problema analogia, nel corso di una singola elaborazione. Questo aspetto viene sottolineato da Mitchell e Hofstadter nel riassumere le potenzialità della rete semantica di COPYCAT e, perciò, nel dare conto del suo *potere rappresentazionale tout court*:

> Poiché il grado di similarità tra due nodi è dipendente dal contesto, i concetti nella rete di slittamento sono *emergenti* piuttosto che definiti esplicitamente. Essi sono *associativi* e *dinamicamente sovrapposti* (qui la sovrapposizione è modellata dai collegamenti) e il loro comportamento che varia nel tempo (attraverso l'attivazione dinamica e il grado di similarità) riflette le proprietà essenziali delle situazioni incontrate. In tal modo i concetti sono in grado di adattarsi (in termini di rilevanza e similarità l'uno con l'altro) a differenti situazioni. Si noti che non stiamo modellando il *learning* nel senso usuale del termine: il programma non mantiene i cambiamenti nella rete di esecuzione in esecuzione, né crea nuovi concetti permanenti; tuttavia, il nostro lavoro implica il *learning* se questo termine viene considerato includere la generalizzazione dall'esperienza [la concettualizzazione] che gli esseri umani mettono in atto nei contesti nuovi. (Mitchell, Hofstadter, 1990, p. 325 [enfasi mia])

Questo passo riassume mirabilmente tutte le caratteristiche che un modello subcognitivo come COPYCAT possiede al fine di simulare un'attività concettuale di tipo *human-like*. Non è l'unico[20]. Tuttavia, esso pone l'attenzione su più di una questione fondamentale. Innanzitutto, crea un collegamento esplicito fra *emergenza*, *associazionismo* e *dinamicità* dei concetti, le tre caratteristiche su cui è basata la natura *fluida* e *creativa* della conoscenza che i modelli sono progettati per simulare dal punto di vista della loro *capacità concettuale*. Anzi è proprio la loro *dinamicità adattiva* in quanto operazione inversa al processo di emergenza a far sì che essi possano essere considerati nei termini della *capacità di modellare l'intensione concettuale* nel descrivere situazioni specifiche, piuttosto che, anche se non necessariamente in contrapposizione, nei termini del possesso di una conoscenza fatta di *liste di tratti in grado di definire lo spazio della estensione concettuale*, la quale, però, sarebbe solo un modo di vedere il sistema a posteriori, cioè a

[20] Ad esempio, per una fonte in italiano in cui si discute diffusamente di questi aspetti del modello si rimanda al già citato Mitchell, Hofstadter (1994).

elaborazione avvenuta. Infatti, solo a esecuzione terminata è possibile descrivere in modo statico la conoscenza effettivamente impiegata e i concetti utilizzati dal sistema.

In secondo luogo, ci viene fatto notare che la teoria del modello fa largo uso della nozione di adattamento, nella quale si riassume e si identifica, in senso generale, il potere rappresentazionale. Al sistema, cioè, verrebbe meno la sua qualifica di *intelligente*, ovvero di modello *effettivo* dei meccanismi del pensiero, se fosse privato della sua capacità adattive costitutive della *correlazione rappresentativa*, attraverso le quali si aggirano gran parte delle obiezioni rivolte all'iconismo rappresentazionale statico di cui molti sistemi di IA simbolica sono stati accusati.

Infine, ma posto come limite empirico, viene sottolineato come il modello non sia dotato di una capacità di *learning*, che peraltro non rientra nei suoi scopi primari, poiché "dimentica" le trasformazioni che il suo *dinamicismo rappresentazionale* ha apportato alla parte concettuale dell'architettura. In tali modificazioni, tuttavia, è lecito vedere (e questo può essere fatto dall'esterno, osservando l'esecuzione e il risultato finale del programma, facendo valere la sua caratteristica di *modello*) le «generalizzazioni dall'esperienza» (*generalization from experience*) che COPYCAT pone in atto nella risoluzione dei problemi che gli vengono sottoposti, testimoniando, dunque, dei meccanismi che sono alla base di questo processo conoscitivo astrattivo.

3.3.2 METACAT e i suoi prolegomeni

Uno sviluppo nella direzione dei problemi lasciati in sospeso da COPYCAT è presente nel sistema che ne costituisce, pur nelle differenze, l'ideale evoluzione: METACAT. Questo programma, come dice il nome, intende porsi a un meta-livello rispetto al suo predecessore, mantenendo intatto il dominio di applicazione (l'alfabeto e i problemi di analogia fra stringhe di lettere) e ponendosi come fine la simulazione di aspetti di *learning* e di capacità *metacognitive*. Una discussione in merito coinvolge i tratti architetturali che lo contraddistinguono e, in via preliminare, i suoi presupposti teorici. Tuttavia, è necessaria una premessa. Molti degli aspetti che sono stati messi in evidenza nella descrizione di COPYCAT caratterizzano anche altri modelli subcognitivi. A partire da qui in avanti, perciò, per evitare ridondanze la descrizione che daremo dei modelli sarà meno particolareggiata e focalizzata sulle differenze più che sulle somiglianze, e ci soffermeremo sulle idee alla base di alcune scelte determinanti dal punto di vista simulativo e, dunque, relative al fenomeno cognitivo indagato. Il criterio esposto all'inizio in merito ad una classificazione per domini di applicazione sarà mantenuto.

In un saggio dal titolo succintamente programmatico e riccamente allusivo, *Prolegomeni ad ogni futuro METACAT* (Hofstadter, 1995b), torna a mostrarsi il motivo kantiano sotteso all'impostazione della ricerca in questo tipo di approccio all'IA. Seppure vada ricordato che il richiamo a Kant è soltanto indicativo delle tematiche trattate, non certo di una metodologia di indagine filosofico critica, coerentemente con la dicitura del titolo Hofstadter indica quale seconda delle caratteristiche

che sembrano «rendere coscienti i cervelli» (Hofstadter 1995b), l'*auto-osservazione*. Questo aspetto non era stato affrontato in COPYCAT, nel quale, invece, trova ampia trattazione e sviluppo quella che considera la prima caratteristica essenziale della «*peculiare organizzazione*» (*ibidem*) che rende i cervelli coscienti: il *possesso concettuale*.

Il richiamo alla capacità di auto-osservazione è un richiamo a Kant nella misura in cui nella sua *Critica della ragion pura* una funzione essenziale del pensiero viene conferita all'"Io penso", unità originaria dell'appercezione, che permette l'unificazione nel giudizio del molteplice empirico attraverso le categorie concettuali. Tuttavia, ciò che va contro uno degli assunti fondamentali dei modelli subcognitivi è l'assenza di una qualche unità di controllo centrale, sulla cui implausibilità psicologica è stato scritto molto e numerose sono state le critiche, in base all'argomento della *reductio ad absurdum*, portate nei confronti di un centro apprensivo delle rappresentazioni interno alla mente, critiche che non è possibile ripercorre in questa sede. Accenniamone, perciò, solo alcuni tratti e riferimenti.

Dennett è stato uno dei più strenui oppositori dell'ipotesi del "teatro cartesiano" della mente, ovvero, quel luogo del pensiero (*cogito*) di cartesiana memoria su cui le rappresentazioni sarebbero *rappresentate* a beneficio dei meccanismi del pensiero (ad esempio, Dennett, 1998). Molti sono stati i luoghi in cui il problema dell'identità dell'io sono stati affrontati. Alcune risposte ai quesiti del problema del sé sono già nei capitoli conclusivi di Hofstadter (1979) in cui, sulla scia della discussione delle conseguenze filosofiche dei teoremi gödeliani in merito alle limitazioni della natura del pensiero, egli propone una possibile spiegazione riduzionista del sé contro le tesi impossibiliste di Lucas (1961) in merito all'effettiva simulatività di questa caratteristica del pensiero in un modello cognitivo implementabile al calcolatore. A queste tematiche è dedicato ampio spazio anche in Hofstadter, Dennett (1981).

Tuttavia, va fatto notare che unità di controllo centrale e teatro cartesiano della mente non sono lo stessa cosa, ma coincidono quanto più si pensa che l'attività del cervello a un certo livello può essere analizzata ipotizzando un qualche tipo di funzionamento attraverso rappresentazioni. Infatti, la sede di impiego di tali rappresentazioni è molto facilmente individuabile proprio nell'unità di controllo centrale del sistema, la quale è certamente una delle componenti fondamentali del calcolatore di von Neumann, e, già antecedentemente, della Macchina di Turing. La nascita di posizioni all'interno dell'IA e delle scienze cognitive in netta opposizione con l'immagine di una mente che agisce in maniera sequenziale, e mono- e centro-diretta, è andata di pari passo, negli scorsi decenni, con la definizione dei tratti più cospicui e fondativi dell'approccio connessionista, per il quale una corretta simulazione dei meccanismi del pensiero non può non affondare le sue radici nella replicazione del funzionamento cerebrale, parallelo, distribuito, auto-organizzato e auto-diretto. Anche da questa prospettiva sono giunte numerose critiche sia al rappresentazionalismo dei sistemi sia alla delega della funzione di controllo ad un'unità apposita dedicata. La non evitabilità di un modulo di questo tipo, in base all'assunto per cui comprendere, spiegare e riprodurre la mente

vuol dire comprendere, spiegare e riprodurre *solo* il cervello, va vista come il segno del fallimento di ogni sistema (con la pretesa di essere) intelligente basato sul simbolismo rappresentazionale.

Tuttavia, non sembra sia possibile liquidare la questione molto facilmente. Le attività *meta*-cognitive non hanno (ancora?) trovato una spiegazione attraverso la costruzione di modelli connessionisti. Il loro darsi sembra legato proprio a quelle capacità auto-osservative coscienti o semicoscienti, la cui descrizione in termini diversi da quelli simbolici sembra ancora del tutto irraggiungibile, o, perlomeno, molto lontana. METACAT, come programma che si pone l'obiettivo di superare le limitazioni di COPYCAT, integra l'architettura di questo con alcune componenti prettamente simboliche, atte a simulare i processi auto-osservativi e più squisitamente astratti tipici del pensiero umano. Gli obiettivi specifici che Hofdstadter ascrive a questa evoluzione di COPYCAT riguardano le possibilità di un sistema il quale possa «*autoesaminarsi*, cosa che permette la nascita di un complesso modello interno di sé, [portatore di] un altissimo grado di autocontrollo e di apertura» (Hofstadter, 1995b, p. 335). Tali obiettivi consistono, dunque, nella implementazione delle *meta*-capacità messe in atto dagli uomini nell'affrontare compiti in domini specifici, quali sono il riconoscimento di un percorso già compiuto, ovvero, generalizzando, il ricordare azioni appena effettuate; la capacità di riconoscere le soluzioni date da altri ad un determinato problema, ovvero la giustificazione di una soluzione già fornita (quanto al grado di salienza, profondità, astrazione, ecc.); un «forte senso "meta-analogico", cioè la capacità di vedere le *analogie tra le analogie*» (*ivi*, p. 341); infine, la capacità di produrre nuove analogie, che coinvolge il «senso estetico» (*ivi*, p. 342) di ogni agente cognitivo, nella misura in cui la creazione di un nuovo problema deve possedere requisiti di profondità e astrazione e candidarsi ad essere riconosciuta come la *variazione sul tema più appropriata*[21].

Il progetto METACAT è stato sviluppato da Jim Marshall[22] sfruttando un'architettura arricchita di tipo FARG. L'obiettivo non è quello di creare un programma alternativo a COPYCAT, bensì costruire un'estensione del modello che incorpori, fra le altre cose e soprattutto, la capacità di *auto-osservazione* summenzionata. Per tale ragione, gli aspetti più significativi di questo modello sono le aggiunte architetturali rispetto alla struttura triadica di COPYCAT, aggiunte che rientrano tutte nella base di conoscenza concettuale del programma. In altri termini, alla consueta Rete di Slittamento sono affiancati altri tre moduli, la cui funzione è differente, ma il contenuto è, generalmente parlando, costituito da concetti. Tale modo di vedere le cose rende esplicito il collegamento fra livello dell'attività meta-cognitiva e apparato simbolico-concettuale del sistema,

[21] Hofstadter è in molti passi esplicito in merito all'importanza (che implica, gioco forza, la loro necessaria considerazione ai fini simulativi) di alcune caratteristiche tanto elusive quanto pervasive dell'attività mentale, come, ad esempio, nelle seguenti righe: «Credo, infatti, che la sensibilità alla *bellezza* e alla sua stretta parente, la *semplicità*, abbia un ruolo centrale nella cognizione di alto livello, e mi aspetto che, via via che le scienze cognitive progrediranno, si arriverà ad ammetterlo con sempre maggiore chiarezza» (Hofstadter, 1995b, p. 342). Si noti la stretta parentela di queste affermazioni con l'opinione diffusa fra i matematici sulla rilevanza di un analogo senso estetico che guidi i passi e le mosse di una dimostrazione matematica.

[22] Si rimanda a Marshall (1999) per un'esposizione completa e a Marshall (2002, 2006) soprattutto per gli aspetti relativi al *self-watching*. Si veda anche Hofstadter, Marshall (1998).

nella misura in cui la simulazione della prima avviene sulla base di una componente concettuale semanticamente forte e referenziale, messa altresì in atto dal programma. Consideriamo ora le tre parti in modo più specifico, evidenziando le funzioni che sono chiamate ad assolvere nel modello, la cui architettura è rappresentata nella figura 3.3.

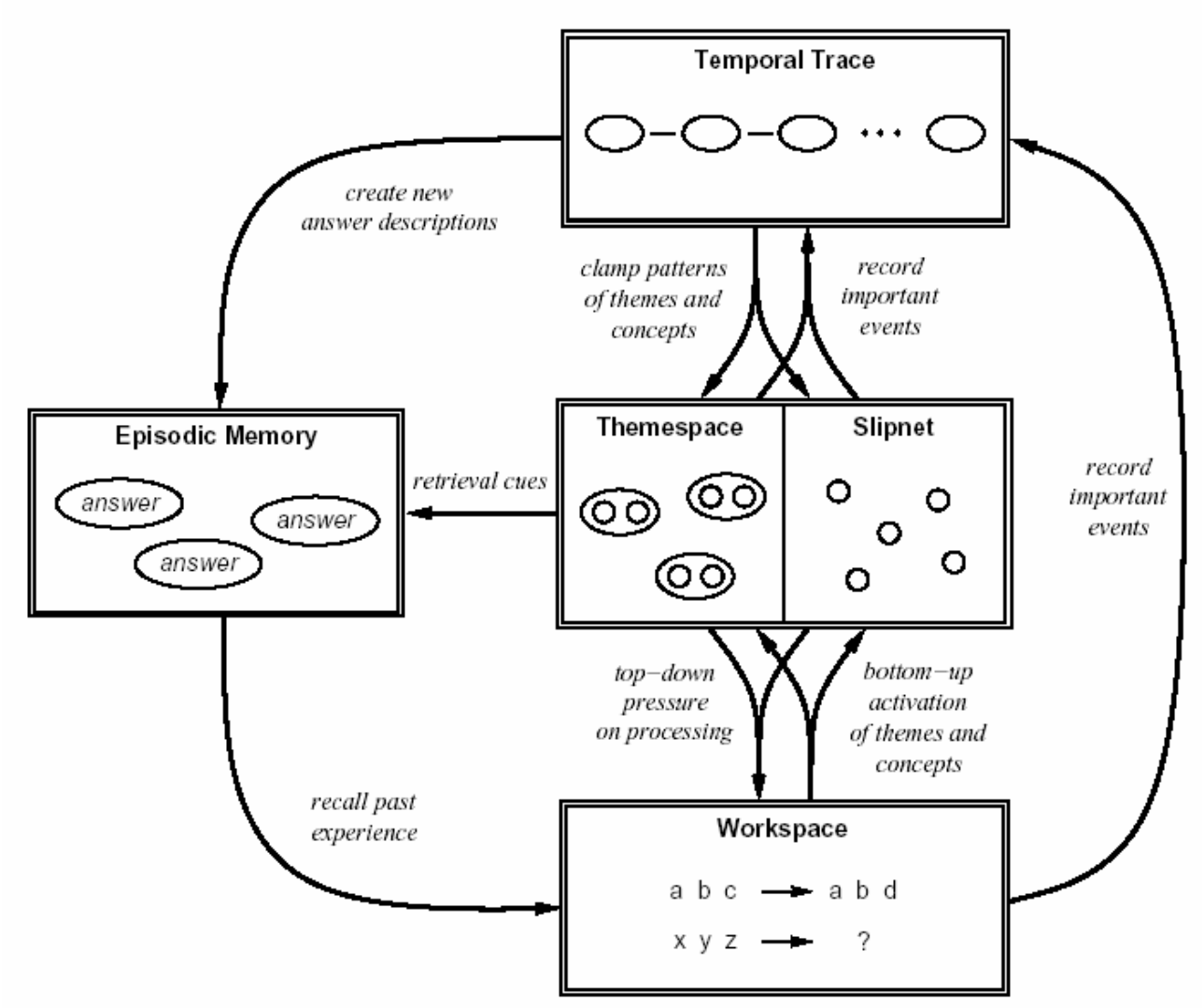

Fig. 3.3 - L'architettura di METACAT (tratto da Marshall, 1999, p. 56)

La componente aggiuntiva più importante è lo Spazio dei Temi. In esso sono contenuti i "temi", particolari coppie di concetti che hanno la seguente forma:

(1) *String-Position: identity*

Come si vede, il primo concetto, denotando una particolare categoria sotto cui un oggetto può ricadere (in questo caso una lettera), esprime una proprietà; il secondo, invece, una relazione, che in questo caso è quella di identità. L'unione di questi due tipi concetti (proprietà + relazione) è, come

nel caso di tutta l'attività elaborativa-rappresentazionale del programma, conseguenza dell'elaborazione stessa, ovvero è METACAT a creare i temi a seconda di ciò che esperisce nello Spazio di Lavoro. Nello Spazio dei Temi sono, perciò, inserite coppie di concetti che «sono in primo luogo e soprattutto strutture rappresentazionali [...]. Ma a certe condizioni, quando fortemente attivati, possono anche esercitare pressioni rilevanti di tipo *top-down* sui processi subcognitivi di METACAT» (Marshall, 1999, p. 57), esattamente come fanno i concetti della rete semantica. Perciò, i temi sono, da una parte, strutture "percepibili" da parte delle microprocedure, perché costruiti dal programma, dall'altra, come i singoli concetti, contribuiscono a guidare e determinare la direzione del processo elaborativo. Il fatto che essi siano costruiti a partire dall'esperienza del programma, attraverso l'attività di elaborazione, li rende controparte esplicita di quel processo di *presa di consapevolezza* della situazione che in COPYCAT non era espresso sotto forma di *pattern* di attivazione della rete semantica:

> Parlando in termini generali, il livello di attivazione di un tema è inteso rappresentare quanto esplicito è il livello di "consapevolezza" di METACAT in merito a una particolare idea nella sua interpretazione corrente di un problema di analogia. Ad ogni istante determinato, molte idee sono presenti *implicitamente* nello strutture dello Spazio di Lavoro che sostanziano i processi di mappatura tra stringhe, ma temi fortemente attivati rappresentano il riconoscimento *esplicito*, da parte del programma, dell'importanza di certe idee. In altri termini, l'attivazione di un tema riflette l'ammontare di "evidenza" che esiste in favore del ritenere che quella particolare idea giochi un ruolo importante nella caratterizzazione della situazione in oggetto (*ivi*, pp. 130-131).

I temi ricoprono, dunque, il ruolo di *trait d'union* per eccellenza fra percezione e cognizione, essendo sia prodotti che, al tempo stesso, guide del processo elaborativo, cioè *da parte di* e *per* il programma. Svolgono nel sistema la funzione essenziale di *portare all'evidenza* ciò che il programma fa, rendendo disponibile questa informazione al modulo percettivo-attivo del programma stesso, che, esattamente come in COPYCAT, è costituito dagli agenti micro-procedurali. In tal modo, i prodotti del pensiero possono diventare oggetto degli stessi meccanismi che li hanno generati, in un *processo di retroazione* che conduce, come già per COPYCAT, alla determinazione di un unico cammino di soluzione. Come si vede nello schema della fig. 3, i temi (coppie di concetti) e i concetti costituiscono il nucleo centrale dell'architettura, determinando l'attività che avviene nello Spazio di Lavoro e al tempo stesso anche il funzionamento degli altri due moduli originali dell'architettura di METACAT[23].

[23] Tutto il processo ovviamente avviene sulla base dell'attività dei codicelli che sono rappresentati dalle linee nere fra le parti dell'architettura. Manca, infatti, in esso il modulo che raccoglie l'elenco delle microprocedure con le relative urgenze. Tuttavia, l'esclusione è presumibilmente dovuta al fatto che esse, a differenza delle altre, sono l'unica componente rappresentazionale puramente procedurale del programma.

Allo stesso modo dei temi, che essendo un prodotto dell'elaborazione del programma, a differenza dei concetti della rete semantica, non sono presenti all'inizio di ogni lancio di METACAT, la Traccia Temporale costituisce un modulo meta-cognitivo, vuoto all'inizio dell'elaborazione, che ne registra in corso gli accadimenti più importanti, come l'attivazione di certi concetti o temi, la costruzione di regole di trasformazione fra le prime due stringhe o fra le seconde due, gli slittamenti (di attivazione) fra i concetti della rete, e così via. Tale modulo, perciò, rappresenta la conoscenza che il programma ha del suo *comportamento*, immagazzinato sotto forma di memoria, a breve termine, della concatenazione temporale delle azioni che METACAT compie. La Traccia Temporale contiene una conoscenza *esplicita*, che, per il fatto di seguire un andamento temporale, può essere considerata *semi-causale*, nel senso che il programma la registra secondo una visione di anteriorità-posteriorità, caratteristica necessaria, anche se non sufficiente, di ogni rapporto causa-effetto. In altri termini, essa esprime la conoscenza *conscia* del programma, cioè il percorso compiuto a livello macroscopico dall'elaborazione intesa in senso globale. Perciò, è attraverso di essa che il programma esercita il grado massimo di *auto-osservazione* e di *auto-controllo* del cammino di soluzione.

Conoscere il proprio comportamento attraverso una successione di azioni macroscopiche permette al programma di evitare il ripetersi di situazioni (*snag-problem*), ma anche di affrontare nello stesso modo positivo problemi, o sotto-problemi, che ha già affrontato in precedenza. Il punto essenziale sta nel fatto che sono sempre la microprocedure a gestire questo tipo di conoscenza, così che si può affermare che il programma ha e non ha allo tempo stesso il medesimo punto di vista sulla situazione in oggetto: lo ha, se si considera che sono sempre le stesse microprocedure in azione nel corso di tutta l'elaborazione; non lo ha, se si tiene conto dei differenti tipi di conoscenza (memorizzata) in gioco. Queste possono essere considerate appartenere ad almeno tre livelli diversi (Marshall, 1999, p. 162): quello subcognitivo, costituito dagli elementi nello Spazio di Lavoro; quello intermedio rappresentato dai temi; quello cognitivo in senso proprio rappresentato dalla memoria temporale delle azioni macroscopiche salienti del proprio comportamento, la cui plausibilità psicologica va rintracciata, tuttavia, più nella funzione di miglioramento della prestazione col procedere dell'elaborazione (e l'inserimento di eventi salienti nella traccia) che nella rappresentazione simbolica delle azioni stesse. Ciò è dovuto al fatto che *ogni* livello è dotato di un'interpretazione simbolica *dall'esterno*, mentre *dall'interno*, pur essendo gli elementi di ogni livello soggetti a uno stesso strumento "percettivo", cioè le microprocedure, *come se* il programma li considerasse dall'esterno, la *simbolicità* differisce da livello a livello in una gerarchia di simboli *omogenei* allo stesso livello ed *eterogenei* fra livelli diversi. La distinzione, interrelata, fra questi ultimi preserva la differenza fra le diverse *qualità simboliche*.

L'architettura di METACAT è, dunque, basata su un sapiente equilibrio fra forme di procedura e forme di memoria, la cui accessibilità ad ampio raggio garantita agli stessi meccanismi procedurali crea quella sorta di «"collasso di livelli" tra i livelli cognitivi e subcognitivi» (*ivi*, p. 163) che

costituisce, come si è visto, una delle caratteristiche principali di ogni modello che voglia simulare capacità percettive di alto livello. La differenziazione delle forme di memoria è portata a un grado ancora più elevato di specificazione in METACAT con il modulo della Memoria Episodica, che raccoglie gli aspetti rilevanti di ogni elaborazione (che giunge a ipotizzare una soluzione) in unità-ricordo riutilizzabili per elaborazioni future. La Memoria Episodica costituisce, dunque, un altro tipo di memoria a lungo termine, insieme ai concetti della rete semantica e all'insieme delle microprocedure, in grado di guidare le elaborazioni future nel momento in cui la ripetizione di un sufficiente numero di tratti di un episodio passato sia nuovamente "attivo", poiché ricreato nello spazio dei temi. In questo modo avviene il recupero dell'informazione e il programma *ricorda* l'episodio già "vissuto", cioè elaborato.

Tale modulo intende modellare la capacità di immagazzinare e richiamare esperienze passate, grazie all'inserimento nella Memoria Episodica delle tracce temporali delle elaborazioni filtrate dei loro componenti meno rilevanti. Il richiamo di episodi passati influisce poi direttamente sull'attività dello Spazio di Lavoro, cioè sulle microprocedure predisposte alla costruzione delle strutture percettive. In definitiva, e in maniera del tutto plausibile psicologicamente, attraverso la simulazione di questa capacità di *reminding* o *recalling*, il programma viene dotato di un forte strumento per l'apprendimento inter-elaborazione. Grazie ad esso, la conoscenza prodotta in determinate circostanze può essere utilizzata nuovamente in circostanze uguali, ovviamente, o anche soltanto simili, quando cioè soltanto alcuni tratti della situazione in corso, sotto forma di temi attivi, producono il superamento della soglia del richiamo di un ricordo passato. Tale superamento può avvenire in differenti circostanze di attivazione, cioè sulla base di insiemi di tratti differenti.

La complessa architettura di METACAT cerca di simulare quasi tutti i differenti tipi di memoria (semantica ed episodica, a lungo termine e a breve termine, dichiarativa, procedurale, locale e distribuita, anche se non subsimbolica), nonché di elaborazione (seriale e parallela), che sono, tipicamente in maniera dualistica, alla base della varie teorie cognitive sulla memoria[24]. Attraverso questi meccanismi il programma è in grado di risolvere problemi di analogia, ma anche di fare ipotesi sul come si è arrivati ad alcune soluzioni. Non è in grado, attraverso le componenti di cui è dotato, di inventare nuove analogie, né tanto meno di produrne di nuove "originali" e "poco banali". Manca, cioè, di alcune capacità che avrebbero incrementato il suo grado di creatività, di cui tuttavia non si può dire del tutto sprovvisto. Come in COPYCAT, infatti, è in grado di dare differenti soluzioni allo stesso problema e di arrivarvi attraverso differenti percorsi di costruzione di rappresentazioni scopo-specifici. D'altra parte, i meccanismi auto-osservativi iscritti nella sua architettura gli permettono di avvalersi di un vero e proprio sistema di retroazione, anche se non è l'obiettivo, bensì il *contesto concettuale attivo* in cui esso deve essere prodotto, a essere progressivamente aggiustato dall'interazione fra i diversi moduli.

[24] Si veda il capitolo 2.

METACAT è chiamato a fornire, come il suo predecessore, un output di due tipi: una soluzione e una regola di trasformazione che *descriva* / *giustifichi* il cambiamento della stringa obiettivo in quella di risposta come analogo a quello fra la stringa sorgente e quella modificata. Nel compiere questo processo, in un'apposita finestra vengono riportate, in linguaggio naturale, le azioni compiute dal programma, una sorta di commento al suo operato. Esso è indipendente dal fatto di essere considerato un commento per altri o un discorso compiuto fra sé e sé, e la sequenza delle operazioni enunciate alla fine viene riassunta in una finestra di commento sempre in linguaggio naturale. Ciò non deve trarre in inganno. Questo programma, come gli altri dedicati all'implementazione della fluidità concettuale, non sono pensati per comprendere e produrre il linguaggio naturale. Il loro utilizzo di espressioni in linguaggio naturale è soltanto un aiuto per l'osservatore esterno, un'interfaccia che facilita la comprensione del comportamento del programma. Infatti, come osserva Marshall in merito alla funzione di creazione di regole di trasformazione, ed è un'osservazione che può essere estesa a tutti i modelli che ricadono all'interno di questo approccio,

> tutta l'informazione che caratterizza unicamente la regola è presente nella sua struttura concettuale sottostante, che è il *solo* livello rappresentazionale che realmente conta. (*ivi*, p. 96 [enfasi mia])

La centralità di questa affermazione risiede nel fatto che essa è ancora una presa di posizione contro una considerazione meramente esteriore dell'attività del programma. In altri termini, nelle discipline simulative l'output e la forma che esso prende attraverso la modalità interfaccia di cui un sistema viene dotato non sono rilevanti ai fini della componente esplicativa, *a meno che* essi non diventino oggetto di rappresentazione per il modello stesso, grazie a una circuitazione circolare dei livelli. Di conseguenza, è evidente come la *questione del giusto livello rappresentazionale* è qualcosa che precede epistemologicamente, in fase di definizione delle restrizioni simulative, la *questione del giusto livello ontologico* del fenomeno simulato. Conseguenza ulteriore, e specifica in merito al programma che stiamo considerando, è che nella regola di trasformazione che viene prodotta ciò che deve essere considerato sono, da una parte, i concetti utilizzati per riempirla, dall'altra, il modo con cui si è arrivati a riempirla *proprio* con quei concetti. Per quanto riguarda il secondo aspetto, esso è soddisfatto dalla descrizione dell'architettura del programma. In merito al primo, occorre dire che la regola altro non è che una maschera, una struttura sintattica con alcuni spazi vuoti che vanno riempiti con i concetti appropriati, la cui individuazione può essere considerata frutto dell'attività *selettiva* ed *emrgente* di elaborazione.

In particolare, la costruzione delle regole in METACAT, in termini molto generali, consiste nel riempimento di una sagoma (*template*) della seguente forma:

Replace _____ *of* _____ *by* _____

Sono i concetti attivati nel corso dell'elaborazione, e filtrati attraverso i meccanismi esaminati in precedenza, a riempire gli spazi vuoti (*slot*). I concetti che esprimono una regola di trasformazione «consistono di una lista arbitrariamente lunga di *clausole di regola*» (*ivi*, p. 89) che descrivono gli oggetti (le lettere e i gruppi di lettere) da un punto di vista interno (*intrinsic clause*) ed esterno (*extrinsic clause*), nel primo caso con riferimento alle caratteristiche specifiche, cioè le *proprietà*, dell'oggetto in questione; nel secondo esprimendo i rapporti con altri oggetti, dunque relativamente alle *relazioni* che l'oggetto considerato ha con altri nell'ambiente percettivo. Inoltre, sono presenti altri *slot* che si riferiscono a nodi categoria o descrittivi o a relazioni presenti nella rete. Ciò che ne deriva alla fine è una lista di proposizioni espresse in un linguaggio formalizzato del primo ordine, che, da una parte, permette il confronto fra regole e tutte le possibili relazioni di accoppiamento e sovrapposizione, e, dall'altra, è indice della natura *puramente simbolica* del livello a cui vengono condotte, dal punto di vista implementativo, le operazioni di messa a confronto a fini analogici.

Non è possibile, per ragioni di spazio, proseguire nell'analisi dettagliata di tutti gli aspetti di questa funzione del programma. Quanto detto dovrebbe aver largamente comprovato come una delle possibili forme di evoluzione e arricchimento del potere di fare analogie passa attraverso la dotazione di moduli architetturali dal forte connotato simbolico, che pure agiscono sulla base dei passi compiuti e dei risultati rappresentazionali raggiunti dalle parti di basso livello del programma. Nondimeno, è la presenza di un forte elemento simbolico a rendere METACAT molto più potente in termini di prestazioni rispetto al suo predecessore, senza privarlo delle capacità di rappresentazione percettiva e di fluidità concettuale che, all'interno dell'approccio subcognitivo, garantiscono la plausibilità psicologica del modello e lo salvaguardano dal ricadere nel problema epistemico della vuotezza di rappresentazioni intese in senso puramente sintattico

3.4 Il mondo dei numeri in successione

3.4.1 SEEK-WHENCE e gli schemi numerici

Lo studio delle successioni numeriche costituisce uno dei punti di partenza delle ricerche hofstadteriane in IA. Uno dei primi modelli progettati, e in parte realizzati, per la comprensione della successioni numeriche, SEEK-WHENCE, aveva come scopo principale l'indagine dei meccanismi cognitivi coinvolti nella estrapolazione degli schemi (*pattern*) di successioni di numeri naturali quali:

(i) 1 1 2 1 2 3 1 2 3 4 1 2 ...

(ii) 1 2 2 3 3 4 4 5 5 6 6 7 ...

Questi sono solo due esempi[25] di successioni la cui analisi e comprensione rientra fra gli obiettivi del modello. Al di là della complessità delle prestazioni presentate e auspicabili da parte del programma, lo scopo generale perseguito da Hofstadter con questo modello è quello di testare alcune delle caratteristiche principali dell'intelligenza (espresse dalla lista fornita nel capitolo precedente), piuttosto che la messa in opera di corpose e ricche abilità matematiche. In altri termini, negando di voler produrre un sistema esperto, Hofstadter così si esprime nel ricostruire retrospettivamente questa esperienza intellettuale:

> è ovvio che è necessaria una *certa* conoscenza dell'ambiente per poter partire, ma avevo la sensazione profonda che l'intelligenza abbia, e *debba* avere, un poderoso nucleo generale astratto, indipendente dalla conoscenza stessa. (Hofstadter, 1995, p. 50)

Tale nucleo generale è espresso poco più oltre da Hofstadter con una lista di caratteristiche che ricalca quella fornita quasi venti anni prima in *Gödel, Escher, Bach* e che viene genericamente da lui definita come «sensibilità per le strutture, che comprende attività del tipo: notare le uguaglianze [...], notare le relazioni semplici [...], notare le analogie [...], imporre la coerenza [...], costruire astrazioni [...], spostare i limiti [...], cercare la bellezza» (*ivi*, p. 57). Ancora una volta si ha una sorta di distinzione fra le capacità dominio-specifiche che sono richieste per la comprensione di un determinato dominio e le capacità generali, applicabili in ogni dominio, che costituiscono gli aspetti essenziali di ogni comportamento definibile come intelligente e oggetto di studio delle discipline simulative. Ne consegue che per indagare tali qualità ancora una volta la scelta di un micro-dominio viene visto come un dispositivo metodologico che serve a concretizzarne la messa in atto, altrimenti soltanto individuabile a livello teorico, e non comprovabile nella prestazione. Altra conseguenza è che il modello non deve necessariamente essere dotato di conoscenze matematiche raffinate, quanto piuttosto deve avere, per così dire, una *conoscenza di tipo percettivo relativamente ai numeri*, quella che si potrebbe anche definire come una *matematica ingenua dei numeri naturali*. Ne deriva, infine, il fatto di risultare ulteriormente rafforzata l'idea che l'intelligenza oggetto di indagine delle discipline simulative, *non può non* presentarsi come processo, e dunque come prestazione, su "materiale concreto", e implicare una componente imprescindibile di percezione, seppure di alto livello.

Il programma SEEK-WHENCE è stato uno dei primi progetti del FARG ad essere realizzato (Meredith, 1986). Anche esso aveva lo scopo di «esplorare questo nuovo universo del non-verbalizzabile, della corrente sottomarina del mentale, del "subcognitivo"» (*ivi*, p. 4).

[25] Un elenco di successioni la cui comprensione costituisce l'obiettivo del programma è reperibile in Hofstadter (1995a, pp. 68-69).

Consideriamone gli aspetti essenziali. In generale, il programma può essere pensato come un modello predittivo, che ricevuti uno alla volta i termini di una successione formula un'ipotesi su quale sia la regola sottostante la successione stessa in modo da poterne indovinare il termine successivo. Da questo punto di vista, l'ipotesi che viene proposta dal programma può essere considerata una *teoria esplicativa* del mondo cui si applica, cioè la successione. Come ogni teoria, essa non è in sé compiuta ed esatta in senso assoluto, bensì sempre perfettibile. Ogni nuovo termine che devia dalla previsione attesa porta al cambiamento dell'ipotesi formulata, così come ogni evento non spiegabile da una qualche teoria scientifica, passati tutti i controlli relativi alla sua "misurazione", impone la revisione della teoria. Il mondo delle successioni numeriche si presta decisamente all'indagine dei meccanismi cognitivi insiti in questo processo di revisione e quindi di *slittamento concettuale contestuale*[26]. Infatti, è sempre possibile che il numero seguente della successione modifichi la regola di produzione valida in precedenza. Così, se le prime cifre di una successione sono: 4, 5, 4, 5, la regola soggiacente può essere espressa dalla seguente proposizione:

Regola 1: "Ripeti i numeri 4 e 5 in questo ordine all'infinito".

Tuttavia, con l'introduzione di una quinta cifra, ad esempio 6, diversa da quella attesa in base all'ipotesi della regola 1, cioè il 4, sono costretto a rivedere la regola di produzione della successione e a trasformarla, ad esempio, nella seguente:

Regola 2: "Aggiungi il numero naturale successore dell'ultimo numero del precedente gruppo ascendente e ricomincia il conteggio da 4".

Un tale tipo di universo permette, dunque, l'indagine dei meccanismi preposti non tanto all'individuazione di complesse funzioni matematiche, quanto alla ricerca di regolarità soggiacenti a un insieme di elementi di natura omogenea, i numeri naturali, la cui relazione reciproca è esprimibile attraverso le semplici funzioni di successore, predecessore, identità, così come già era stato per il dominio delle lettere dell'alfabeto, con la differenza che qui vengono considerati, per *default*, i puntini alla fine della successione come la possibilità di una sua infinita continuazione, e, di conseguenza, di una sempre possibile modificazione della regola di produzione soggiacente. Un modello capace di muoversi in questo dominio non deve avere conoscenze matematiche più che elementari, ma deve essere in grado di cogliere la presenza di schemi ripetitivi ed esprimerli in una regola e, tuttavia, di saper modificare le proprie convinzioni acquisite, cioè essere *flessibile*. La

[26] Interessante da questo punto di vista il dialogo hofstadteriano che descrive i diversi atteggiamenti assunti da alcuni interlocutori all'atto di ipotizzare il modo in cui una successione può continuare al variare, in maniera incrementale, delle cifre disponibili, cioè "scoperte". Di fatto essi corrispondono alle *euristiche* attraverso cui ci si muove nello spazio virtualmente infinito delle possibili successioni, a partire da un frammento di lunghezza qualsivoglia di cifre, purché non minore di 1, al fine di rintracciare lo schema sotteso alla successione che idealmente continua il frammento dato. Si veda Hofstadter (1983b, pp. 11-34).

regola deve essere inferita dal confronto dei vari elementi (singole cifre o gruppi di cifre) grazie alla scoperta dei tratti invarianti fra distinti gruppi di cifre. Come è facile vedere, anche in questo caso come in quelli di COPYCAT e METACAT, la regola di produzione della successione, il cui ritrovamento è lo scopo del programma, è individuata *per via analogica* attraverso il reperimento di *invarianti concettuali* fra i segmenti della successione.

Perciò, come sottolineato da Hofstadter (1982, pp. 5-6)[27], anche in questo caso si tratta di dotare il programma di una capacità adeguata di rappresentazione della situazione attraverso un opportuno repertorio concettuale, affinché possa formulare le possibili spiegazioni alternative della successione e scegliere quella migliore secondo un'intuizione estetico-economica. Le due assunzioni di principio su cui essa si basa e che divengono le euristiche del programma sono individuabili in uno humeano principio di uniformità della natura e in un'opzione verso la *semplicità* della spiegazione: «in qualche senso, dunque, le spiegazioni "semplici" sono quelle più corte possibili» (*ivi*, p. 5), cioè quelle che contengono il minor numero di elementi possibile dotati della minore complessità.

È interessante notare come nell'inedito citato siano già presenti molti degli elementi che caratterizzeranno, quantomeno dal punto di vista della rappresentazione della conoscenza, l'impostazione algoritmica di tutti i modelli subcognitivi dello stesso tipo successivi a SEEK-WHENCE, il primo ad essere effettivamente implementato. I bersagli polemici di Hofstadter sono espliciti. Il più immediato sono i modelli per l'estrapolazione di sequenze – tra i quali è maggiormente noto quello proposto da Simon e Kotovsky (1963) – che fanno uso di tecniche di ricerca quasi *brute-force* nell'albero che costituisce lo spazio delle soluzioni possibili delle spiegazioni di un successione data. Tale procedimento consiste nell'applicazione ricorsiva di un ristretto numero di funzioni ai numeri della successione, funzioni che incorporano una dose approfondita di conoscenza matematica da utilizzare, in maniera abbastanza tipica nell'IA del primo decennio, come un "setaccio euristico" nello spazio problemico. Non c'è, insomma, l'applicazione selettiva di euristiche metodologiche non matematiche, bensì è l'impiego di un insieme di funzioni matematiche a costituire l'*euristica*, che è definibile come tale solo perché è una restrizione rispetto all'insieme di tutte le funzioni applicabili. Tali modelli non colgono effettivamente secondo Hofstadter i veri meccanismi sottesi all'analisi umana delle sequenze numeriche, basati secondo lui maggiormente su più fondamentali capacità di scoperta dei *pattern*, più o meno astratti, presenti in una situazione. In definitiva, anche in questo caso sembra trattarsi di qualcosa di molto simile a *primitive relazionali*, in particolare per quanto riguarda le relazioni di ordine e di identità (o similarità sotto certi aspetti).

Il secondo bersaglio polemico sono le tecniche di rappresentazione delle conoscenza in voga nell'IA partire dagli anni settanta, in particolare le «reti per le descrizioni strutturate» (*ivi*, p. 6) di

[27] Si tratta del primo documento sull'argomento, inedito e reperibile nell'archivio del Center fo Research on Concepts and Cognition dell'Indiana University.

Winston (1975b), attraverso cui un programma costruisce la propria rappresentazione di un universo epistemico a partire da esempi, accusate di essere «completamente statiche o dichiarative» (*ibidem*). A questo *modus operandi* rappresentazionale Hofstadter oppone in maniera molto nitida la sua concezione di «conoscenza di cui è dotato un programma» relativamente al dominio delle successioni numeriche:

> [...] noi vogliamo che la rappresentazione di una data sequenza assomigli a ciò che *percepiamo* come lo schema (*pattern*) stesso, piuttosto che un programma che produca lo schema. Il compito allora diviene quello di immettere informazione procedurale in una maschera statica senza trasformare una struttura trasparente in un programma opaco. (*ibidem*, [enfasi mia])

Il passo è illuminante poiché mostra come Hofstadter, pur utilizzando un linguaggio ancora tipico di una certa impostazione di IA legata alla disputa fra rappresentazioni della conoscenza dichiarative e procedurali, ipotizzi un trascendimento di questa dicotomia con l'introduzione di un modo di rappresentazione della conoscenza, per così dire, *simbolicamente procedurale*, dove il termine "procedurale" non indica meramente l'implementazione operativo-algoritmica delle parti attive del programma, distinte da dati simbolici staticamente strutturati e detentori esclusivi della conoscenza *significativa* del programma. Una rivisitazione della contrapposizione dichiarativo-procedurale appare implicare per Hofstadter la possibilità di avere *simboli procedurali dotati di significato*, e questo tanto più nella misura in cui una simile capacità viene attribuita al pensiero umano, dotato, per fare un esempio attinente al modello, della possibilità di esprimere una parte di una successione attraverso l'applicazione di una regola e una successione illimitata attraverso una regola più complessa. In queste idee è forse rintracciabile la riproposizione di un antico problema filosofico, quello della pensabilità dell'infinito in intensione più che in estensione, la quale mette in luce nella proposta hofstadteriana echi aristotelici e leibniziani.

Il superamento della contrapposizione procedurale dichiarativo non consiste, dunque, nel rigettare il simbolico, ma nel modificarlo, poiché, in ultima analisi, *tutto dipende dalla giusta rappresentazione simbolica scelta dal programma*, e, perciò, dalle capacità di cui viene dotato a questo fine, che è anche quello di poter manipolare le sue rappresentazioni. Non è difficile vedere come la proposta di un simbolismo procedurale, che sia tale per il programma, da una parte schiude la via ai modelli seguenti in grado di mettere in pratica il processo di percezione di alto livello in maniera sempre più raffinata; dall'altra, costituisce l'aspetto complementare del problema della giusta notazione per la descrizione di una successione, visto che «una descrizione di una sequenza nella notazione di SEEK-WHENCE dovrebbe generalizzarsi (o slittare) facilmente a descrizioni di sequenze in modo da costituire l'"essere come"» (*ivi*, p. 5), ovvero la trasposizione analogica di una successione in un'altra.

Possibili soluzioni al problema di simulare i meccanismi cognitivi del processo di generalizzazione sembrano costituire lo scopo principale della creazione di un programma come SEEK-WHENCE, soprattutto poiché il dominio in cui si muove il modello, come si è detto, è, da una parte, omogeneo – gli elementi base sono i numeri naturali la cui manipolazione appare sia facilmente implementabile su un calcolatore sia plausibilmente, dal punto di vista psicologico, ascrivibile a un apposito modulo cognitivo (indipendentemente dal sua radicamento a livello cerebrale) – e, dall'altra, passibile di un prolungamento all'infinito, caratteristica che vincola l'individuazione di una regola, diversamente che nei modelli precedentemente analizzati, alla sua effettiva applicazione ciclica. Questo aspetto è proprio ciò che viene garantito dalla capacità cognitiva di generalizzazione, la quale può arrivare a essere considerata l'anello di congiunzione fra i processi di categorizzazione e quelli di creazione di analogie, via la nozione più generale di *pattern recognition*, o di estrapolazione di strutture. Infatti, come afferma Hofstadter, proponendo una nuova lista di tratti che ricorda da vicino, ma in maniera più operativa, procedurale, quella individuata per denotare il nucleo dell'attività "intelligente":

> [...] nel pensiero umano la generalizzazione è molto, molto più ricca della semplice sostituzione di costanti con variabili. Generalizzare significa avere la capacità di riconfigurare internamente un'idea, così:
> - muovendone avanti e indietro i confini interni;
> - scambiando i componenti o spostando sottostrutture da un livello a un altro;
> - fondendo due sottostrutture in una, o dividendone una in due;
> - allungando o accorciando un componente dato;
> - aggiungendo nuovi componenti o nuovi livelli di struttura;
> - sostituendo un concetto con uno molto simile;
> - verificando il risultato di inversioni su vari livelli concettuali. (Hofstadter, 1995a, p. 92)

Torniamo ora al modello e consideriamone, per ragioni di spazio, solo i tratti essenziali. SEEK-WHENCE procede segmentando le successioni alla ricerca di *raggruppamenti significativi* di numeri, sia in merito delle loro caratteristiche intrinseche (ad esempio, percependo (1 2 3 4) come gruppo ascendente nella successione (i)), sia per quanto riguarda gli aspetti contestuali (ad esempio, vedendo (3 3) o (5 5) come gruppi di cifre simili all'interno di una catena di coppie di cifre identiche nella successione (ii)). Per ottenere un'opportuna strutturazione della sequenza SEEK-WHENCE si avvale di un meccanismo complesso di rappresentazione della situazione percepita. Innanzitutto è dotato di un repertorio concettuale, «una rete di *concetti primitivi*, [...] fissati, in quanto relazioni di livello basico» (Meredith, p. 12 [enfasi mia]), tutti rappresentati come procedure, seppure non relativamente a un livello informatico-implementativo, bensì a livello

cognitivo esplicito come una sorta di *regole di produzione di numeri espresse in forma simbolica*[28]. A partire da questi sono costruiti i concetti complessi corrispondenti a strutture articolate, sulla base delle cui parti vengono compiuti confronti e stabilite relazioni di accoppiamento e somiglianza: «questa rappresentazione "complessa", strutturale, di concetti permetterà l'uso di similarità strutturali come "collegamenti virtuali" nel sistema. In altri termini, possiamo correlare due concetti notando le somiglianza nelle loro strutture e/o nei blocchi di cui le strutture sono costruite, piuttosto che guardando semplicemente alla loro lista di attributi» (*ibidem*). Tale struttura concettuale permette, attraverso la manipolazione dei suoi componenti, la revisione della descrizione dello schema, qualora una nuova cifra, cioè un nuovo fatto introdotto nell'universo, induca la necessità di questa trasformazione, che avviene sulla base dei concetti contenuti nella MLT del programma.

Da un punto di vista macroscopico, la rappresentazione della conoscenza del programma si articola su tre livelli, che attraverso la combinazione dei concetti primitivi (procedurali) in concetti complessi, cerca di ottenere una scansione completa della successione. Nessun numero viene lasciato fuori, cioè non considerato dal programma, e grazie a tale scansione viene formulata un'ipotesi della regola soggiacente, ipotesi che, se non confermata, è rigettata in vista di una più appropriata. I tre livelli della rappresentazione del programma sono distinti a seconda della complessità delle strutture di dati che li caratterizzano. A grandi linee, si può dire che esiste uno spazio percettivo, in cui sono presenti i valori della successione, percepiti come strutture atomiche e indissolubili (chiamate "*glint*"). Questi vengono raggruppati in strutture di livello immediatamente superiore (i "*glom*"). Tutto ciò avviene nel livello percettivo inferiore. Infine, nel livello intermedio della rappresentazione del programma sono presenti strutture aggregate più complesse (gli "*gnoth*"), sui quali il programma opera attivamente, scambiandone pezzi e ristrutturandoli sulla base dei concetti platonici del livello superiore, che costituiscono la conoscenza permanente del programma costituita dai tipi ideali, cioè i concetti sulla base dei quali le strutture sono formate. Il livello intermedio, perciò, è il luogo in cui spinte dal basso e dall'alto si confrontano: verso il basso per via del processo di creazione di schemi (*template*) mediatori fra il livello inferiore e quello intermedio, verso l'alto nel processo di produzione di ipotesi, reso possibile dall'interazione fra i concetti platonici e le strutture costruite nel livello intermedio.

Solo per fare un esempio, prendiamo brevemente in considerazione la successione (i):

1 1 2 1 2 3 1 2 3 4 1 2 3 ...

[28] I concetti primitivi di SEEK-WHENCE sono funzioni a uno o più argomenti, che restituiscono come valori numeri naturali interi non negativi. Ad esempio, Countup (*x*), che restituisce il valore di *x*, poi *x*+1, poi *x*+2 e che sta ad indicare la funzione di successore; C-group (val, *n*), che restituisce il valore della variabile ripetuto per *n* volte; Cycle (arglist), che restituisce un gruppo di numeri in maniera ciclica. Tali primitive, come si vede, costituiscono una conoscenza di tipo procedurale, nel senso che il programma *conosce* la sequenza in oggetto, cioè se la rappresenta, attraverso regole di produzione di alcune sue parti.

Intuitivamente, è possibile considerare una buona prestazione del programma una segmentazione del tipo:

(1) (1 2) (1 2 3) (1 2 3 4) (1 2 3 ...

Una tale rappresentazione corrobora la previsione che il numero successivo della sequenza sia 4. La costruzione di tali raggruppamenti avviene sulla base delle spinte della rete concettuale. Finché SEEK-WHENCE non è in grado di considerare nella sua visione della successione tutte le cifre, continuerà ad esplorare la successione. Una volta presi in considerazione tutti i numeri e fornitane una spiegazione funzionale, microprocedure apposite costruiranno legami fra le varie strutture per arrivare a una visione unitaria – in genere una concatenazione di funzioni primitive nidificate del tipo visto in precedenza – che l'ipotesi esprime in forma proposizionale come regola della successione.

Al di là dei meccanismi effettivi del programma, che sono, per la verità, piuttosto complicati e di non sempre facile accostamento con i moduli funzionali cognitivi (MBT e MLT) tipici dei sistemi presi in considerazione precedentemente, vanno sottolineati alcuni aspetti rilevanti di questo modello. In primo luogo, esso, pur nell'opacità cognitiva delle sue strutture di dati[29], anticipa gli sviluppi dei modelli futuri realizzati all'interno dell'approccio subcognitivo, sviluppi che troveranno una formulazione più omogenea e più referenzialmente cognitiva, per quanto riguarda la terminologia impiegata nel descrivere i modelli. Ciò, d'altra parte, rispecchia una tendenza generale delle scienze cognitive degli ultimi decenni, cioè il progressivo abbandono di una terminologia tecnico-ingegneristica nella descrizione di modelli esplicativi/simulativi dei meccanismi del pensiero. Anche in questo va visto uno dei tratti distintivi della trasformazione e dell'inclusione dell'IA nella più generale scienza cognitiva, un'evoluzione che, oltre ad allargare positivamente il campo di indagine, ha, come contropartita, generato dispute, contrasti, fraintendimenti.

In secondo luogo, in SEEK-WHENCE è già modellata quella capacità di slittamento concettuale che avviene grazie alla particolare struttura della rete semantica (di slittamento). Nella riformulazione della ipotesi, infatti, in caso di non predittività, la rappresentazione strutturata della sequenza, presente nel livello intermedio, è passibile di cambiamento attraverso lo spostamento di attenzione dal concetto o dai concetti motivanti l'ipotesi/regola della successione ad un altro o ad altri, sulla base di pressioni contestuali e del meccanismo parallelo di indagine del materiale percettivo, così come si è visto in COPYCAT. Tale processo prende la forma, a livello percettivo, di differenti raggruppamenti della cifre della successione, sulla base dell'attivazione di nuovi concetti. Il punto centrale è che tutta questa attività ha luogo progressivamente, senza una distinzione netta di fasi antecedenti di rappresentazione e conseguenti di formulazione di ipotesi, da un massimo di

[29] Il livello concettuale è chiamato "Platoplasma", quello intermedio "Socratoplasma", dove il suffisso –plasma richiama la metafora, già menzionata, dell'elaborazione come processo "enzimatico" attuato dalle microprocedure. Una descrizione dettagliata delle componenti funzionali del modello è in Meredith (1986, pp. 48-85).

stocasticità, nell'assenza di strutture percepite, a un massimo di determinismo, che consiste nella creazione della regola esplicativa-produttiva della sequenza. Questo schema di elaborazione sarà poi generalizzato a tutti i modelli dell'approccio subcognitivo.

Anche riguardo a SEEK-WHENCE, proprio in ragione del dominio in cui esso agisce, acquistano fondamentale importanza le relazioni di uguaglianza e di successione fra gli elementi, ovvero, la formazione, nella terminologia insiemistica, di partizioni (gruppi di elementi e, successivamente, raggruppamenti di gruppi) secondo classi di equivalenza e classi ordinate. Questo era stato anche il caso del modello proposto da Simon e Kotovsky, di cui SEEK-WHENCE mostra di essere una ripresa, soprattutto nella misura in cui anche tale modello era volto all'indagine dei meccanismi che guidano il processo induttivo di acquisizione concettuale[30]. Tuttavia, seppure essi ammettano che i soggetti umani coinvolti in esperimenti con le successioni compiono in maniera simultanea l'analisi in termini di relazioni primitive della successione e il processo di formulazione dell'ipotesi che spiega la successione, affermano che il loro programma «separa le due fasi dell'attività di *problem-solving* – individuazione della periodicità e descrizione dello schema (*pattern*) – in maniera più netta rispetto ai soggetti umani» (Kotovsky, Simon, 1973, p. 410).

La svolta messa in atto in SEEK-WHENCE è quella di non operare i due processi separatamente: «i processi di costruzione e revisione [produzione progressiva della regola della sequenza] procedono in parallelo con quelli di analisi» (Meredith, 1986, p. 131), attraverso le primitive relazionali. Dunque, questo modello «prende in considerazione l'ipotesi corrente alla luce dell'evidenza del nuovo termine e tenta di cambiare la forma dell'ipotesi per includere il nuovo termine» (*ibidem*). SEEK-WHENCE introduce, perciò, per la prima volta un modo differente di guardare ai procedimenti induttivi in domini ben strutturati, ponendo l'accento sull'idea che anche i problemi "più logici" in cui l'IA si era cimentata nei decenni precedenti, ad esempio in un universo come quello matematico dei numeri naturali, trovano una soluzione migliore se il modello cognitivo si basa su euristiche alternative all'applicazione esclusiva di procedimenti basati sulla logica dei predicati[31] o sulla separazione netta in sottoproblemi di tipo differente per facilitare l'applicazione in sequenza di insiemi di funzioni, come nel modello di Simon e Kotovski. L'utilizzo del meccanismo della "scansione parallela a schiera", descritto nel precedente capitolo, permette la maggiore flessibilità del modello di Meredith. Poiché anche in esso è presente un'operazione di descrizione della successione basata sull'applicazione di funzioni seppure molto semplici (le microprocedure esplorative), l'elaborazione parallela da stati stocastici a stati deterministici e l'interazione fra processi di analisi e produzione di ipotesi sono da considerarsi le vere euristiche su cui si basano le potenzialità di SEEK-WHENCE.

[30] «La periodicità è determinata dal cogliere relazioni I e N (*identity* e *next*)» (Kotovsky, Simon, 1973, p. 410), cioè identità e successività, ovvero, come interpreta Meredith in termini più deboli, ma adatti a descrivere la natura dell'operazione di rappresentazione fluida e dinamica, secondo relazioni basate sulle *nozioni euristiche* di «uguaglianza e successore» (Meredith, 1986, p. 131).

[31] Si consideri, ad esempio, in merito a questo approccio il programma SPARC/E da Dietterich e Michalski (1985), di poco precedente la realizzazione di SEEK-WHENCE.

L'utilizzo di quella che si può definire "euristica parallelistica" è un punto di svolta nell'IA e nelle discipline simulative alla metà degli anni ottanta del secolo scorso. Basta ricordare che la pubblicazione che dà l'avvio a questo nuovo tipo di approccio, *Parallel distribuited processing* a cura di Rumelhart e McClelland, risale al 1986, lo stesso anno in cui SEEK-WHENCE viene sviluppato nella sua prima versione definitiva. Tuttavia, l'approccio subcognitivo condivide con il connessionismo solo alcuni, e non la totalità dei, principi metodologici, non spingendosi fino alla negazione del simbolico in direzione di una simulazione diretta, ancorché semplificata, delle reti neuronali, senza considerare, peraltro, le differenze metodologiche e di principio che attraversano la ricerca connessionista considerata da un punto di vista globale. SEEK-WHENCE ha bisogno dei dati espressi in forma simbolica, ancorché svincolata da eccessive restrizioni al suo livello più basso, proprio nella misura in cui aspira a essere un modello delle capacità subcognitive basate su primitive relazionali della percezione, possibilmente estendibili, nelle intenzioni degli autori, oltre il suo dominio specifico di applicazione. Parlando dei modelli in fase di progettazione da parte del FARG alla metà degli anni ottanta, Meredith dice:

> «quando tutti i sistemi saranno completati, saremo auspicabilmente in grado di astrarre le caratteristiche comuni, indipendenti dal dominio che sono utili in senso generale. Questa è una strategia "alto rischio, alto guadagno. Noi speriamo che funzioni.» (*ivi*, p. 149)

Va notato, infine, come la differenziazione in SEEK-WHENCE di tre livelli di strutture di dati sublimerà poi, in altri modelli sviluppati nell'ambito del FARG, in architetture dai connotati più psicologistici. In genere si avranno due fondamentali ambienti della conoscenza, corrispondenti alla MBT, o spazio di lavoro, e alla MLT, o rete concettuale permanente, che possono essere considerati non più in maniera gerarchica, ma in rapporto di mutua influenza, come si è visto. I tre livelli di SEEK-WHENCE, quello percettivo di base, quello delle strutture e quello dei concetti, appaiono nei modelli successivi come risultato dell'elaborazione emergente e lasciano libero il campo da troppo rigide separazioni gerarchiche fra tipi di dati, seppure in esso si attua già un superamento rispetto alle tecniche precedenti di rappresentazione della conoscenza. Tale superamento è funzionale alla simulazione di caratteristiche del pensiero che nella loro generalità, in questa prospettiva, entrano in gioco anche, e non solo, in domini come quello della matematica, in cui una certa percezione estetica delle strutture favorisce l'acquisizione di nuove scoperte e la concettualizzazione di regolarità prima celate attraverso procedimenti fallibili, ma ristrutturabili, riprogrammabili.

3.4.2 SEQSEE e le nuove strategie auto-osservative

SEEK-WHENCE è un modello manchevole sotto diversi aspetti rispetto a quelli che lo hanno seguito. Non è presente in esso, ad esempio, la possibilità di immagazzinare nuovi concetti a

elaborazione terminata, i quali sono risultato dell'elaborazione stessa. Ancora, manca di alcune delle caratteristiche dinamiche della rete concettuale, o di un meccanismo auto-osservativo complesso, come quello visto in METACAT.

Superare alcune di queste mancanze è l'obbiettivo dello sviluppo di un nuovo modello, attualmente in fase di costruzione ad opera di Abhijit Mahabal: SEQSEE[32]. Il programma è in grado di determinare quale sarà il numero successivo di una sequenza data in input. Tuttavia, non è ancora provvisto di una memoria a lungo termine, cioè di una repertorio concettuale stabile, che gli permetta di cogliere le regole soggiacenti a sequenze del tipo:

(iii) 2 15 16 1 2 3 8 9 10 11 5 6 7 8 9 11 12 13 14 15 ...

o anche:

(iv) 1 1 2 3 1 2 2 3 1 2 3 3 1 1 2 3 ...

Questa incapacità sono interessanti perché ci dicono qualcosa sugli aspetti architetturali che, invece, mettono in condizione il programma di trovare il termine successivo di sequenze come:

(v) 1 2 3 4 5 ...

o

(vi) 1 1 2 3 1 2 2 3 4 1 ...

Ricordiamo che l'ottica di costruzione del sistema mira a simulare la formulazione di conoscenze per via induttiva. Per tale ragione non si parla di una conoscenza assoluta, né di determinazione di regole assolute. Ciascuna scelta sulla definizione della regola di una successione è, in linea di principio, aleatoria e non necessaria. Di conseguenza ogni volta che una successione viene continuata, il nuovo termine può rimettere in gioco la precedente regola soggiacente alla successione stessa. Tenuto conto di questo fatto, è ovvio come ogni successione spinga a vedere la propria regola di formazione, la quale, come si è detto, è dal punto di vista di un soggetto umano quella più semplice ed economica, se si accetta il principio di economia della spiegazione scientifica, la moderna versione dei rasoio di Occam. Perciò, se prendiamo come esempio le successioni appena tracciate, possiamo vedere che la (iii) può essere segmentata nel seguente modo:

[32] Alcune notizie relative a SEQSEE sono disponibili sul seguente blog: http://seqsee.blogspot.com/, cui si rimanda per una discussione aperta di alcuni degli aspetti progressivamente implementati nel programma, come ad esempio la sua capacitè di *self-watching*.

(iii) (2) (15 16) (1 2 3) (8 9 10 11) (5 6 7 8 9) (11 12 13 14 15 ...

e la regola che la spiega è qualcosa del tipo: “aumenta di una unità il gruppo di successione seguente quello considerato”. Tuttavia, possono darsi altre formulazioni maggiormente descrittive e, dunque, più informative. La (iv) è una successione periodica o ciclica ed è frammentabile nella maniera seguente:

(iv) ((1 1) 2 3) (1 (2 2) 3) (1 2 (3 3)) ((1 1) 2 3) ...

La regola che la esprime potrebbe essere: “duplica in successione tutti i numeri del gruppo (1 2 3) e ripeti all’infinito la sequenza”. Ciò non toglie che l’n-esimo numero della successione potrebbe mutare anche questa regola, deviando dallo standard, cancellando la periodicità e autorizzando a differenti raggruppamenti delle cifre. La (v) non ha bisogno di spiegazioni. La regola più immediata per un soggetto umano che la descrive è analoga alla definizione per induzione dell’insieme dei numeri naturali. La (vi) è solo poco meno intuitiva e si può suddividere così:

(vi) (1 1 2 3) (1 2 2 3 4) 1 ...

La regola di questa successione potrebbe essere così espressa: “duplica il terz’ultimo numero di ogni gruppo di successione e aggiungi il numero successivo di ogni gruppo”.

SEQSEE ha a disposizione un repertorio di circa quindici codicelli o microprocedure per poter effettuare tutte le operazioni di legame, corrispondenza per identità o somiglianza, e raggruppamento. Tuttavia, la mancanza (per ora) di un repertorio concettuale diminuisce notevolmente la sua capacità di afferrare le regole più astratte delle successioni. Di conseguenza, sembra si possa affermare che questa è un’altra dimostrazione del fatto che le *regolarità* basate su concetti astratti richiedono una modellizzazione esplicita della conoscenza del programma, effettuata attraverso l’implementazione di una rete semantica di simboli espliciti e ricollegabili come *tipi* all’attività delle microprocedure. Infatti, mentre è possibile che un’istanza di un numero venga considerata uguale a un’istanza di un altro numero, attraverso una semplice microprocedura di accoppiamento che agisce direttamente nello Spazio di Lavoro sugli elementi della successione presenti, non è possibile fare confronti fra complessi o composizioni di azioni delle microprocedure, se esse non trovano un riscontro nell’attivazione di un determinato concetto o repertorio concettuale, che in qualche modo, stando all’architettura di base su cui questi modelli sono costruiti, è necessario per *cristallizzare* porzioni di elaborazione che producono risultati in questo senso emergenti, cioè non direttamente implicati nell’esecuzione di una *singola* microprocedura. In definitiva, l’assenza dei concetti limita il potere astrattivo del sistema. Inoltre, la mancanza di un

corredo di concetti impedisce al programma di vedere lo *snag-problem* presente nella sequenza (iv), che ripete lo stesso periodo all'infinito.

È interessante notare che uno dei nuovi moduli architetturali presenti nel programma è il così detto "Flusso di Pensieri" (*stream of thought*), che richiama la Traccia Temporale di METACAT in vista della modellizzazione di un processo di auto-osservazione alternativo a quello del programma di Marshall. Infatti, mentre la Traccia Temporale conservava sullo stesso piano tutti gli eventi salienti di un'elaborazione, il Flusso di Pensieri conserva i "pensieri correnti" del programma (gli ultimi dieci nella versione attuale) e, dunque, non deve andare alla ricerca di sovrapposizioni con pensieri prodotti in tutto il corso dell'elaborazione. Questa caratteristica del modello da un parte ottimizza le sue risorse computazionali, dall'altra è intesa cogliere un aspetto psicologicamente plausibile del pensiero umano, ovvero quello per cui, al di là di ciò che noi possiamo tenere a mente in merito agli input di un problema che ci viene sottoposto, sui quali eventualmente possiamo sempre ritornare attraverso il supporto grafico (cartaceo o di qualsiasi altra natura) per mezzo del quale ci viene somministrato il quesito (in questo caso un inizio di successione), non sono molti i pensieri presenti alla nostra attenzione conscia e, in più, sono soggetti al processo di decadimento della memoria a breve termine. Il Flusso di Pensieri di SEQSEE, perciò, può essere visto introdurre surrettiziamente una distinzione fra quello che negli altri modelli era considerata la MBT, cioè lo Spazio di Lavoro, che in SEQSEE costituisce un modulo a parte dell'architettura, e una MBT vera e propria, il cui oggetto sono i pensieri correnti che superano una certa soglia, anche se non elevata come quella degli eventi conservati "a lungo", cioè per tutta l'elaborazione, nella Traccia Temporale, e che è soggetta a un forte processo di decadimento.

Tuttavia, proprio questa ultima debolezza del Flusso di Pensieri sottolinea la necessità di un immagazzinamento degli eventi nella memoria in maniera più stabile che ancora manca a SEQSEE e che potrebbe essere oggetto di un modulo a parte, il quale d'altra parte condivida la caratteristica già presente in METACAT di rendere disponibile l'informazione lì contenuta all'analisi delle microprocedure proprio allo stesso modo, cioè come se fosse *allo stesso livello*, di quella presente nello Spazio di Lavoro. Tale collasso dei differenti livelli, cioè dei *modi funzionali*, in cui è implementata la conoscenza del programma costituisce uno degli obiettivi futuri di questo modello[33].

Infine, la scelta di un dominio come quello delle successioni numeriche, che, come si è detto, presuppone per definizione la possibilità dell'infinito, cioè della continuazione infinita della sequenza, è alla radice di una potenzialità peculiare riservata ai modelli che operano in domini di questo tipo. Se, infatti, dal punto di vista effettivo, l'eventualità di una continuazione all'infinito si traduce nella possibilità di un input iniziale indeterminato per quanto riguarda il numero delle cifre manifeste della successione all'avvio dell'elaborazione – si possono avere, infatti, input iniziali di

[33] Abhijit Mahabal, comunicazione personale.

quattro cifre, o di sei o di dodici, e così via – il caso più interessante diventa quello dell'input mono-cifra, come ad esempio il seguente:

(vii) 1 ...

In una situazione di questo tipo, viste le caratteristiche del programma, i risultati dati da SEQSEE potrebbero costituire un'apprezzabile fonte di dati e una buona metodologia per vagliarne le risorse creative, una volta che si accettino per buone le cifre di volta in volta proposte dal modello senza rigettarle come "sbagliate"[34], e per testare l'architettura computazionale sperimentata nella simulazione.

3.4.3 SEEK WELL: la matematica come musica

Lo studio dei processi di estrapolazione di strutture cominciato con SEEK-WHENCE ha avuto negli ultimi anni un ulteriore sviluppo, grazie alla contaminazione con un altro settore delle scienze cognitive: la cognizione musicale (*music cognition*)[35].

Uno dei modi di guardare alle successioni numeriche, infatti, può essere quello di considerarle isomorfe a partiture musicali in cui il numero esprime l'altezza della nota sul pentagramma, cioè una particolare nota di un'ottava, una volta fissata a 1 la nota che esprime la chiave della melodia. È possibile pensare a un micro-dominio che costituisca un campo d'azione per un modello di cognizione musicale se compiamo alcune semplificazioni rispetto alle possibilità riservate dalla normale notazione musicale, ad esempio prendendo note di uguale durata, che differiscono solo per altezza (sul pentagramma: DO, MI, LA, ecc.), collegate a posizioni metriche prefissate e suonate una alla volta. Larson (1997) ha proposto un dominio di questo tipo, implementabile in maniera piuttosto semplice su un calcolatore e isomorfo alle successioni numeriche così che l'obiettivo principale, effettivamente realizzabile, di un modello sia quello di "indovinare" l'aspettativa melodica creata da un certo numero di note fornite in input, ovvero, in altri termini, di suggerire la nota successiva della parte di melodia fornita in partenza.

L'aspetto interessante è ancora una volta legato al dominio scelto. Esso è costruito per implementare la teoria delle "forze musicali" proposta dallo stesso Larson (1993) per spiegare il modo in cui si produce in un ascoltatore umano l'aspettativa nei confronti di una determinata sequenza melodica *nel corso* della sua riproduzione. Larson individua tre forze, istituendo un'analogia fra il moto nello spazio fisico e il moto percepito dell'avanzare della linea melodica: la *gravità*, ovvero la tendenza delle note a ridiscendere al "piano" rappresentato dalla nota tonica; il

[34] Ricordiamo che anche in questo caso, come per SEEK-WHENCE, è il programmatore ad accettare o rigettare le cifre proposte dal programma sulla base della *sua* ipotesi corrente.
[35] Per un approfondimento, oltre agli scritti di Larson citati in seguito, si rimanda a Lerdhal, Jackendoff (1983), Lerdahl (2001), Narmour (1992) e Margulis (2005).

magnetismo, fra note che godono di una diversa misura di stabilità; l'*inerzia*, come disposizione dello schema musicale a proseguire in maniera uniforme, ovvero, si potrebbe dire, lungo gli stessi binari. Per fare questo Larson distingue fra livelli superficiali e strutturali della melodia e fra note di riferimento (quelle della scala utilizzata) e note obiettivo (quelle dell'accordo basato sulla tonica).

Larson ha messo alla prova la sua teoria attraverso una serie di esperimenti su soggetti umani[36]. Al di là dei dettagli tecnici musicali e dei gradi di conferma riscontrati nei risultati sperimentali, che pure costituiscono in questo approccio simulativo parte integrante della metodologia standard di implementazione di un modello computazionale, la scelta di questo dominio appare interessante anche da un punto di vista teorico più generale. In effetti, Larson definisce il microdominio da lui scelto "creativo", sottolineando un fattore spesso implicito nel processo di sviluppo che porta all'implementazione di questi modelli, ovvero, la relazione asimmetrica di dipendenza fra processi creativi di pensiero e dominio scelto. In altri termini, non solo la scelta di un dominio ristretto ("micro") è funzionale all'implementazione di un modello liberandolo dai rischi dell'esplosione combinatoria, ma permette anche ai soggetti umani, come dimostrano gli esperimenti compiuti in relazione a questo e ad altri modelli, di muoversi creativamente al suo interno, vanificando la possibilità di impiego di metodi deterministici (spesso basati sulla forza bruta), cui pure gli esseri umani a volte ricorrono, e spingendoli a impiegare operazioni mentali creative data la non esplicitabilità dei processi di pensiero (non coscienti) messi in atto. Questa dal fatto che i processi sono basati su *elementi percepiti in prima istanza come semplici*, circostanza che viene ritenuta imprescindibile e vincolante in fase di implementazione del modello.

Ancora una volta la scelta di un dominio ristretto mostra la sua importanza per l'approccio subcognitivo allo studio della mente. Le teorie di Larson sono diventate parte integrante dello sviluppo di un modello denominato SEEK WELL, dedicato all'estrapolazione di strutture melodiche al fine di catturare il processo di aspettativa melodica. Eric Nichols sta sviluppando[37] il programma, impiegando l'architettura tipica del FARG, ovvero costituita dall'interazione fra uno spazio di lavoro, una memoria concettuale che contiene conoscenza relativa alla musica tradizionale occidentale, e un insieme di microprocedure[38] dedicate alle azioni semplici di collegamento e raggruppamento fra note e gruppi di note (come era stato in SEEK-WHENCE fra numeri e gruppi di numeri) e di applicazione dei concetti contenuti nella rete semantica. Infine, è interessante notare come l'integrazione fra le varie parti sia riservata a un modulo di controllo superiore

[36] Si rimanda a Larson (1997) per una rassegna dei lavori che riportano i resoconti di tali esperimenti.

[37] Comunicazione personale. Il modello computazionale deve essere ancora implementato. Ciò potrà comportare alcune varianti allo schema originale, fatto più che tipico dovuto all'inscindibile vincolo pratico-applicativo cui sottostà la ricerca teorica in questo campo.

[38] Un aspetto terminologico interessante è il fatto che nel modello SEEK WELL le microprocedure sono chiamate "lavoratori" (*workers*) piuttosto che "codicelli" (*codelets*). Questo serve a distinguere una caratteristica del modello computazionale considerato dal punto di vista teorico dalla sue implementazione al calcolatore. Tuttavia, il termine "lavoratori", che suggerisce una *connotazione costruttiva* come tratto fondamentale della microprocedura, sottolinea anche la differenza tra questi modelli e quelli tradizionalmente basati su una prospettiva multi-agente, nei quali la caratteristica principale degli agenti è quella di essere visti come sub-unità attive del programma che interagiscono attraverso uno *scambio informazionale*. Cfr. ancora Hewitt (1977).

(*metacontroller*), che gestisce il *loop* centrale dell'elaborazione e che in qualche maniera sembra travisare l'impostazione totalmente emergente dell'elaborazione data a questi modelli[39].

La valutazione di nuove note alla sequenza, e l'aggiunta di note ipotizzate dal programma, è un processo che in SEEK WELL deve considerarsi analogo a quello compiuto da SEEK-WHENCE nelle successioni. Non è azzardato ritenere che la differenza nei dettagli implementativi, se ci sarà, dipenderà, in questo caso, più dalla proposta di nuovi moduli e tecniche implementative del modello che da questioni riguardanti il dominio scelto. In ciò va vista una volta di più quella aspirazione alla generalizzabilità dei meccanismi cognitivi sperimentati nei modelli simulativi, meccanismi che soprattutto nel caso dei programma analizzati in questa sezione sono ben sintetizzati dalle seguenti parole di Hofstadter:

> Io sono convinto che la percezione di strutture, l'estrapolazione e la generalizzazione siano il punto fondamentale della creatività e che si possa arrivare a capire questi processi cognitivi fondamentali *solo* modellandoli in microambienti il più possibile ristretti e progettati con la massima attenzione. (Hofstadter, 1995a, p. 100)

3.5. Il mondo reale a tavolino

In una situazione come quella rappresentata da alcuni oggetti sopra un tavolo, pur considerandone un numero limitato in maniera conforme alle sue dimensioni finite, sono possibili infinite combinazioni nella disposizione degli oggetti medesimi. Se consideriamo il tavolo come lo spazio fisico finito tra due individui uno di fronte all'altro, i lati ai quali stanno gli occupanti possono essere considerati le rispettive aree di influenza e gli oggetti disposti su una parte come componenti un insieme collegato all'occupante di quella parte. Ora chiediamoci: è possibile che a ogni azione di un occupante sui suoi oggetti corrisponda un'identica azione dell'altro sui rispettivi oggetti?

Questa è la domanda da cui scaturisce l'ideazione di un altro modello computazionale che si avvale dell'architettura sperimentata dall'approccio subcognitivo del FARG: TABLETOP. Il dominio specifico di questo programma è, appunto, una porzione di mondo reale costituita da un tavolino e da due occupanti, uno dei quali è manovrato dal programma medesimo, che giocano il gioco del "fare la stessa cosa". In particolare, l'azione consentita è quella di indicare un oggetto, una suppellettile, da parte del primo occupante dalla sua parte. Il secondo risponderà con l'indicazione di un oggetto che ricopra, dalla propria parte, il *medesimo ruolo* nel momento in cui l'obiettivo è di compiere la *stessa azione*. Se il secondo occupante è impersonato dal programma, l'architettura del modello su cui questo si basa dovrà incorporare una serie di moduli che lo mettano

[39] Ma si vedranno altre eccezioni nel seguito.

in grado di compiere la medesima azione, indipendentemente dal fatto che il risultato sia o meno l'indicare uno stesso oggetto. Infatti, ai due lati del tavolo potrebbero esserci oggetti diversi o uno dei due potrebbe essere totalmente sprovvisto di oggetti.

Lo sviluppo del modello computazionale TABLETOP (French, 1995; French, Hofstadter, 1991; Hofstadter, French, 1992) intende portare l'architettura di base tipica dei modelli FARG nel mondo reale, o, meglio, nella particolare situazione del mondo reale appena descritta. Tuttavia, è necessaria una precisazione. Il mondo reale simulato nell'ambiente in cui opera il programma è estremamente idealizzato. Come è ovvio, soltanto un numero esiguo di oggetti è implementato nella rete semantica che rappresenta la conoscenza del programma. Inoltre, questi oggetti, tutti comuni suppellettili che potrebbero trovarsi sopra un tavolo, come piatti, bicchieri o posate, non hanno parti, dal punto di vista del programma, né sono considerati per la loro forma o la loro costituzione. Ma allora, ci si chiederà, in che cosa questo modello differisce dai suoi precedenti, che pure agiscono su oggetti semplici, "elementari" o "atomici", nel senso che individuati solo in quanto istanze di tipi concettuali in una relazione di identità istanza-tipo?

Per rispondere a questa domanda occorre considerare l'architettura complessiva di TABLETOP, sottolineandone le differenze con il modello con il quale il richiamo è più diretto: COPYCAT. La prima macroscopica differenza risiede proprio nella scelta del dominio e nelle conseguenza che questa scelta comporta. Infatti, mentre in COPYCAT non era rilevante la distanza fisica fra le lettere, essa diventa una variabile fondamentale in TABLETOP e uno degli effettivi punti di contatto, in quanto caratteristica condivisa simulata, fra il dominio idealizzato su cui è in grado di operare il modello e il mondo reale. In quest'ottica, la distanza fisica fra gli oggetti concorre alla formazione di gruppi, tanto quanto la loro prossimità semantica implementata nella rete concettuale (French, 1995, p. 42). Quest'ultima, d'altro canto, è dotata di una serie di accorgimenti atti a cogliere le complesse relazioni di inclusione categoriale che caratterizzano il modo in cui un essere umano considera gli oggetti del mondo reale: «in TABLETOP, una singola categoria è spesso associata con un certo numero di differenti categorie sovraordinate» (*ibidem*). Per tale ragione, le categorie sono definite come "indistinte" (*blurry*), e, dunque, non determinate *a priori*, ma emergenti in quanto risultato dell'elaborazione. E ciò in misura ancora maggiore nel modello della Mitchell, nel quale l'implementazione dei concetti per i tipi di lettere non era soggetta a concatenazioni gerarchiche categoriali.

La rete concettuale di TABLETOP è in grado di modellare una situazione del mondo reale, relativamente al suo dominio di applicazione, in maniera psicologicamente più plausibile, mettendo in atto il tentativo di implementare una serie di teorie sui concetti quali quelle proposte, ad esempio, da Rosch (1976; anche in Rosch, Lloyd, 1978), in merito alla categorizzazione e ai livelli categoriali impliciti nella organizzazione concettuale operata da un essere umano nei confronti del suo ambiente. I livelli categoriali utilizzati in TABLETOP sono tre: uno di base, che comprende le categorie degli oggetti (ad es., "piatto", "coltello"); uno intermedio, in cui sono presenti sia le

categorie cui ricondurre le categorie base (ad es., “posate”) sia le relazioni fra queste ultime (ad es., “più grande di”, “vicino a”); un ultimo livello di concetti più astratti, sovraordinati rispetto a quelli di secondo livello ed esprimenti relazioni fra relazioni (ad es., come era già in COPYCAT, “opposto” in quanto meta-relazione che descrive il rapporto fra “destra” e “sinistra”). Per modellare questa complessa capacità di attribuzione categoriale la rete semantica di TABLETOP ha tre tipi di collegamenti fra i nodi (French, 1995, pp. 62-63):

1) collegamenti ISA (“è un”) in numero maggiore che in COPYCAT, tipici delle reti semantiche tradizionali *á la* Quillian (1968), che esprimono le relazioni fra categorie e istanze particolari;
2) collegamenti etichettati (*labeled*), che, come si è visto in METACAT, sono retti da un nodo congiunto al collegamento fra due nodi, che esprime ordinariamente una meta-relazione (ad es., “opposto”);
3) collegamenti *has-member* (“ha come membro appartenente”), che costituiscono l’inversa della relazione di inclusione rappresentata dai collegamenti ISA.

La rete concettuale di TABLETOP ha diversi aspetti interessanti, connessi con le sue caratteristiche. Innanzitutto essa modella un’organizzazione delle conoscenza basata sulla prossimità categoriale. È possibile esprimere, infatti, l’appartenenza di due concetti alla stessa categoria attraverso due legami ISA dai nodi concetti al nodo che rappresenta la categoria nella quale rientrano. In questo modo, viene modellata la rappresentazione della “prossimità concettuale generalizzata”, che ha luogo sia fra concetti di livello astratto, come era per le reti semantiche dei modelli visti in precedenza, sia per concetti che esprimono categorie di oggetti del mondo reale (ad es., “forchetta” o “bicchiere”), rendendo possibile, inoltre, l’implementazione di relazioni categoriali che sono generalmente considerate conseguenza dell’apprendimento. Tuttavia, da questo punto di vista, TABLETOP non è un modello di *learning* diversamente dal modo in cui lo sono gli altri modelli che abbiamo esaminato[40]. I concetti degli oggetti del mondo reale, infatti, sono inseriti nella rete dal programmatore e la loro formazione *ex novo* non dipende, né deriva, dall’elaborazione. Anche nel caso di TABLETOP la capacità di *learning* è limitata alle trasformazioni *intra*-elaborazione della sua rete concettuale, le quali vengono azzerate alla fine di ogni lancio del programma, una caratteristica ben poco plausibile dal punto di vista cognitivo umano.

[40] Se si eccettua il fatto che un generatore di numeri naturali permette ai modelli che agiscono nel mondo delle successioni, particolarmente in SEQSEE, di “comprendere”, cioè di avere a che fare con, numeri sempre più grandi e non contenuti nella memoria concettuale. Il programma conosce la successione dei numeri naturale, attraverso l’implementazione, come euristica sul dominio, della funzione che li produce e basata sul principio di induzione. Il sistema, dunque, non conosce tutti i numeri, ma la regola per produrre ognuno.

In secondo luogo, la rete è dinamica nel senso che la lunghezza dei suoi collegamenti muta con il variare dell'attivazione del noto che etichetta il legame. Ad esempio, se viene attivato il nodo "opposto", la lunghezza del collegamento fra "destra" e "sinistra", come di tutti i collegamenti cui esso è connesso, diminuisce, favorendo lo slittamento concettuale, cioè il passaggio di attivazione da uno all'altro dei due nodi sotto-ordinati.

Veniamo così alla terza caratteristica della rete, che riguarda la funzione che determina la diffusione dell'attivazione attraverso i nodi. Infatti, la possibilità di un slittamento concettuale, sul quale si basa l'analogia e che consiste, di fatto, nello spostamento del punto di vista dal quale il sistema considera la situazione, è conseguenza della dinamica di diffusione dell'attivazione nella rete, la quale viene calcolata attraverso una formula ben precisa[41]. Tralasciando i dettagli matematici, va notato che la prospettiva teorica in cui è costruita la formula che calcola la diffusione dell'attivazione è intesa cogliere gli aspetti della rete che modellano la conoscenza delle relazioni categoriali secondo legami di associazione concettuale. Perciò, a fronte del fatto che esistono molti percorsi per calcolare la distanza fra due nodi, ne consegue che ognuno di questi percorsi concorre ad aumentare la quantità di attivazione che si diffonde da un nodo verso quelli che gli sono direttamente o indirettamente collegati. Inoltre, poiché, come già era per COPYCAT e METACAT, nodi con un grado maggiore di astrattezza, cioè nel caso di TABLETOP nodi che rappresentano categorie sovra-ordinate, sono dotati di un processo più lento di decadimento dell'attivazione, si ha una diffusione tanto maggiore e più sostenuta nel tempo verso i concetti sottoposti quanto più forte è il collegamento fra nodi superiori e nodi inferiori. Infatti, i legami ISA sono indice di relazione categoriale e la presenza di molti legami ISA che mettono in collegamento due nodi con i medesimi nodi sottoposti è segno di prossimità categoriale, la quale, attraverso i legami ISA stessi, è in questo modo causa di un maggiore e più immediato passaggio di attivazione.

Nella rete semantica di TABLETOP i concetti appaiono essere rappresentati in maniera più approfondita rispetto a quelle dei modelli che li hanno preceduti. In questo, fattore determinate è la scelta del (micro-)dominio di applicazione del modello. In particolare, i vari tipi di collegamento fra i nodi, unitamente alla struttura della rete in generale, sono in grado di rappresentare i due principali aspetti della conoscenza semantica concettuale. Essi costituiscono i due estremi di un unico spettro che, generalmente, va dal concreto all'astratto e sono il risultato di due processi distinti, la categorizzazione e la concettualizzazione, intendendo col primo l'attività di coagulazione dei dati dell'esperienza intorno a punti di attrazione considerati come unità primitive e inscindibili dal punto di vista dell'attività percettiva, e con il secondo il risultato del processo di formazione dei concetti più astratti o più complessi, solo apparentemente, vista l'architettura del modello, attribuibile *in toto* a dinamiche interne al pensiero e scisso, o più distante, dall'esperienza. Nella prospettiva di TABLETOP si deve ovviamente parlare di spettro, sia perché si tratta di conoscenza rappresentata e

[41] L'analogia che viene istituita con questa funzione di calcolo chiama in causa la formula attraverso cui viene calcolato il passaggio di corrente elettrica in un circuito in cui sono presenti delle resistenze. Per i dettagli tecnici si rimanda a French (1995, p. 61).

non di simulazione del processo, almeno per quanto riguarda la categorizzazione, sia perché, in ogni caso, queste due capacità sono considerate frutto dell'applicazione dei medesimi meccanismi già molte volte nominati di percezione di alto livello, che *mediano* fra conoscenza posseduta e dati percepiti.

Infine, dato l'utilizzo di tecniche, e perfino di una terminologia, molto affini a quella delle reti connessioniste, occorre chiarire in quali aspetti le due impostazioni differiscono. Innanzitutto, il più vistoso consiste nel fatto che i nodi della rete semantica, diversamente rispetto a quelli della maggior parte delle reti connessioniste tipica, sono tutti interpretabili semanticamente, sono cioè *simboli dotati di significato*, seppure, in questo come negli altri modelli presentati, essi intendono implementare una ben determinata teoria, quella dei "simboli attivi", su cui ritorneremo nel prossimo e conclusivo capitolo. Basti dire, per ora, che, secondo questa prospettiva, ogni nodo della rete rappresenta sì un simbolo, tuttavia non soltanto attraverso una semplice relazione di riferimento corrispondenziale, come nelle semantiche formalizzate tradizionali (*á la* Tarsky). Ciò che la rete, considerata nel complesso delle sue caratteristiche funzionali, rappresenta è la possibilità e la misura delle azioni che un simbolo è in grado di provocare nell'elaborazione globale del sistema, ovvero, dal punto di vista del modello, la sua *funzione suggestiva*, di motore teorico-ideale-semantico del processo di pensiero, attraverso la modificazione dell'attività e del tipo di agenti subcognitivi coinvolti in un determinato momento del processo stesso.

Secondariamente, e da un punto di vista più tecnico, mentre, come sottolinea French, in una rete connessionista sono i pesi degli archi a determinare la misura della quantità di attivazione, istituendosi così una relazione diretta fra distanza e peso nella rete (a un peso maggiore corrisponde un'attivazione maggiore nel nodo verso cui l'attivazione si propaga), nella rete di TABLETOP, al contrario, «la distanza può essere considerata la reciproca del peso», e di conseguenza «la quantità di attivazione diffusa è (di fatto) inversamente proporzionale alla distanza fra i concetti [...]. Così, più due concetti sono prossimi nella Rete di Slittamento (cioè, più corte sono le lunghezze fra i loro collegamenti), più grande è la quantità di attivazione diffusa dall'uno all'altro» (*ivi*, p. 59). E, come abbiamo visto, l'attivazione di un nodo etichetta (di un legame) causa l'accorciamento del legame stesso, diminuendo la distanza fra i due nodi che collega.

Finora abbiamo visto come la scelta di un dominio che ricalcasse una situazione del mondo reale, ancorché solo parzialmente, e in particolare per quanto riguarda le simulazioni di relazioni di distanza spaziale fra oggetti in uno spazio (de-)limitato e di relazioni fra alcune proprietà degli oggetti stessi attraverso la modellazione semantica dei rapporti di gerarchia categoriale, abbia determinato le peculiarità della rete semantica di TABLETOP. L'influenza del dominio, d'altra parte, è visibile anche in altri aspetti di dettaglio dell'architettura, la quale nei suoi tratti fondamentali è costruita sullo schema generale dei modelli dell'approccio subcognitivo visti fino a questo momento. Infatti, altre parti di essa sono lo Spazio di Lavoro e il modulo delle

microprocedure. Queste ultime vengono impiegate probabilisticamente sulla base delle pressioni derivanti dall'attivazione dei nodi nella rete semantica, che, a loro volta, costituiscono una componente della funzione di valutazione delle strutture costruite nello Spazio di Lavoro. Anche in TABLETOP è presente, infine, il meccanismo di auto-osservazione svolto dalla variabile temperatura, che monitora, *e allo stesso tempo* concorre a determinare, la quantità di andamento stocastico presente nell'elaborazione del programma.

Le microprocedure sono anche in questo caso chiamate a esplorare la situazione corrente e a formare raggruppamenti e collegamenti fra gruppi, che possono essere considerati una sorta di meta-gruppi. Le due più importanti relazioni coinvolte in questo processo, conformemente al dominio scelto, sono quelle di prossimità spaziale e di vicinanza (inclusione e appartenenza) categoriale. Di conseguenza, sulla base di questo duplice tipo di relazioni le microprocedure principali sono i "cercatori di gruppi", i "cercatori dei vicini" di un dato oggetto considerato, i "cercatori delle parti terminali" dei gruppi, oltre a quelli classici volti alla ricerca delle corrispondenze fra due oggetti identici.

Il compito del programma è quello di risolvere un ben determinato problema analogico. Deciso un oggetto iniziale, che viene indicato da una freccia, da parte dell'utente umano, TABLETOP deve trovare un oggetto analogo a quello indicato dalla sua parte del tavolo o, comunque, dalla parte opposta a quella dell'oggetto input. Tuttavia, questo non è un vincolo e possono essere indicati come risposta anche oggetti sullo stesso lato del tavolo in cui sta quello iniziale. Ciò è indice ancora una volta di come i vincoli imposti alla ricerca della soluzione da parte del programma sono solo probabilistici e non deterministici in senso assoluto, anche se la tendenza del programma è quella di andare verso una quantità sempre maggiore di determinismo corrispondente all'acquisizione di un punto di vista proprio, definito e univoco. Forti pressioni concettuali possono, in questo tipo di architettura, portare a qualsiasi soluzione che rientri nell'ambito dello spazio percettivo e a diverse elaborazioni col medesimo input possono corrispondere risposte diverse. In questo modo, si è inteso cogliere, come si è già sottolineato in precedenza, la caratteristica della capacità semantica umana, cioè dell'impiego del suo bagaglio epistemico, di essere non deterministica, bensì polivalente, dal punto di vista del confronto di differenti prestazioni e non all'interno della singola prestazione. Si consideri come esempio immediatamente illuminante di questa capacità (o limite di capacità) quel fenomeno molto conosciuto di percezione di-esclusiva che è il cubo di Necker[42], dove medesimi elementi concettuali possono entrare a far parte di una diversa concettualizzazione dello stesso input percettivo. In fondo, il cubo di Necker è pur sempre un cubo, in qualunque modo lo si guardi.

L'assunzione di fondo che guida la strutturazione dello spazio percettivo da parte di TABLETOP è che l'impossibilità per i processi attentivi di un'organizzazione percettiva conscia concettualmente bifocale è caratteristica dell'applicazione dei concetti a ogni livello. Il programma, perciò, procede

[42] Per un'interessante rassegna e una discussione di numerose illusioni percettive si rimanda al sito: http://www.michaelbach.de/ot/index.html

impiegando come euristica una funzione che calcola la salienza (*salience*) degli oggetti e dei gruppi di oggetti privilegiando di volta in volta *una sola interpretazione* fra le altre dello spazio percettivo, costituito dall'insieme degli oggetti e dei loro rapporti spaziali possibili nello Spazio di Lavoro. Tale funzione euristica si basa su molteplici fattori, quali ovviamente l'attivazione del concetto corrispondente all'oggetto, ma anche la posizione esterna dell'oggetto in un gruppo e la sua corrispondenza con altri oggetti. Inoltre se ad essere preso in considerazione è un gruppo, la sua salienza è data dalla presenza di più oggetti uguali, dalla grandezza del gruppo e, in maniera decisiva, dall'appartenenza degli oggetti (tutti o alcuni) a una categoria sovraordinata comune.

In termini complessivi, tale funzione di valutazione permette la strutturazione dello spazio percettivo secondo un criterio non casuale. In tal senso la funzione euristica diminuisce drasticamente il numero dei raggruppamenti possibili da tenere in considerazione, abbassando considerevolmente il dispendio computazionale che sarebbe richiesto da una ricerca compiuta attraverso un algoritmo di forza bruta. Le corrispondenze, create e valutate secondo il grado di salienza con criteri analoghi a quelli degli oggetti e dei gruppi[43] nella strutturazione dello spazio percettivo, permettono lo slittamento concettuale, che favorisce a catena la riorganizzazione continua dello spazio percettivo fino ad avere poche e alternative strutture in competizione. Mentre nei modelli precedenti presi in esame solo una visione della situazione era permessa e le altre perdenti erano di volta in volta distrutte, TABLETOP è dotato di una funzione, la Visione del Mondo (*Worldview*)[44] che permette il mantenimento delle visioni alternative, perché, se è vero che solo un focus attentivo cosciente viene considerato possibile nei processi di pensiero, è anche evidente che «noi possiamo oscillare avanti e indietro tra due interpretazioni della situazione senza problemi. Questo avviene presumibilmente perché manteniamo una rappresentazione in qualche modo attiva (sebbene sotto la soglia dell'attenzione cosciente) della seconda raffigurazione nei nostri cervelli» (*ivi*, p. 70). Si può dire che l'idea che regola questo processo è che noi non distruggiamo ciò che abbiamo comunque percepito, anche se non vi prestiamo attenzione. La Visione del Mondo, perciò, rappresentata nell'interfaccia del programma da collegamenti continui di contro a quelli tratteggiati che stanno per le visioni alternative, «costituisce un insieme di corrispondenze non-contraddittorie di oggetti e gruppi di oggetti gli uni sugli altri» (*ibidem*), mentre la presenza di rappresentazioni alternative conferisce al programma la possibilità di operare anche attraverso l'impiego del controfattuale nella costruzione progressiva della visione definitiva.

Facciamo qualche esempio. Se ho un bicchiere da una parte e un bicchiere dall'altra, la corrispondenza è univoca e la soluzione del problema di analogia è banale: la risposta consisterà nel bicchiere non indicato all'inizio dall'utente. Se, invece, si hanno da una parte, in questo ordine, due forchette, una tazza e due coltelli e dall'altra, in questo ordine, due forchette, un piatto, due coltelli e

[43] A ben vedere una corrispondenza è la stessa cosa di un gruppo solo che la prima mette in relazioni oggetti e gruppi su parti opposte del piano, il secondo riunisce oggetti dallo stesso lato.
[44] Nella traduzione di Hofstadter, French (1995b) il termine "*worldview*" viene reso con "vista globale". Ci è sembrata più appropriata, a fini esplicativi del modello, una traduzione che mantenesse maggiore aderenza all'originale.

più lontana una tazza solitaria, indicando la tazza della prima serie, il programma sarà forzato a scegliere il piatto fra coltelli e forchette dall'altra parte, piuttosto che la tazza solitaria (interpretazione alternativa), sulla base delle spinte della rete concettuale a considerare i gruppi forchette-oggetto-coltello dotati di una salienza superiore e dunque in grado di causare una corrispondenza fra tazza iniziale e piatto finale. In altri termini, il contesto percettivo influenza la percezione della situazione attraverso le pressioni della rete concettuale e la sua funzione di slittamento (nell'esempio dal nodo "*tazza*" al nodo "*piatto*") favorisce questa interpretazione sulle altre[45].

Non è possibile proseguire oltre in questa sede l'analisi di casi e di elaborazioni effettivamente eseguite dal programma. Esso dimostra di agire in maniera molto simile a quella di soggetti umani cui gli stessi problemi di analogia sono stati sottoposti e ciò anche in problemi dotati di un grosso quantitativo di "rumore informazionale" nell'input percettivo, quale può essere costituito ad esempio da oggetti sparsi sul tavolo e collocati al centro del piano piuttosto che agli estremi. In questi casi il programma si trova spiazzato nelle fasi iniziali dell'elaborazione, non potendo attuare facilmente relazioni di corrispondenza fra opposti lati del tavolo, tendenza che fa parte del suo bagaglio epistemico "innato", cioè immesso dal programmatore direttamente nelle funzioni delle microprocedure. La tendenza a trovare corrispondenze fra insiemi di oggetti, infatti, guida l'elaborazione fin dalle prime fasi. La rappresentazione della situazione, infatti, è compiuta dal programma in una maniera che French definisce "gestaltica" (*ivi*, p. 65). TABLETOP è programmato per considerare prima i gruppi, con l'attivazione immediata del nodo corrispondente "*gruppo*" ogniqualvolta individua un certo numero di oggetti spazialmente vicini. Passa poi all'analisi dei loro componenti. Tale processo è preferito a quello contrario di costruzione di un gruppo a partire dalle componenti.

Il procedimento percettivo *top down* non impedisce a TABLETOP di favorire una costruzione coerente del punto di vista, anzi, come nell'esempio visto prima delle forchette e dei coltelli, le corrispondenze che crea tendono in qualche modo a formare una visione unitaria in cui tutti gli elementi di un gruppo trovano corrispondenza negli elementi dell'altro gruppo. Come fa notare French, questo è in qualche maniera interrelato con la nozione di sistematicità proposta dalla Gentner (1983), a supporto del suo modello cognitivo SME, per spiegare la tendenza alla coerenza dei soggetti nell'operare corrispondenze fra elementi di due insiemi di elementi posti in rapporto analogico di tipo globale secondo un principio di biunivocità *strutturale profonda* (cioè, astratta). Tale modo di vedere le cose attraverso l'architettura proposta da French e in larga parte condivisa dagli altri modelli sembra potersi estendere alla capacità percettiva in generale in quanto obiettivo esplicito di questo particolare approccio simulativo, il quale, ricordiamolo, vede una sovrapposizione quasi perfetta fra fenomeno percettivo e processo di creazione di analogie.

[45] Per una serie di esempi che mostrano le sfaccettature del programma e che denotano la sua "personalità" si rimanda a French (1995, pp. 113-149).

Infine, l'elaborazione basata sulla competizione fra interpretazioni controfattuali pone in evidenza un problema delle reti concettuali che caratterizzano le architetture FARG, definito da French «il problema dei singoli nodi con attivazioni molteplici»[46] (*ivi*, p. 82), che, brevemente, si può riassumere in questi termini. Se uno stesso concetto viene attivato più volte in differenti contesti, come può accadere nel caso di un concetto che rappresenta un relazione spaziale come "a destra di", mi trovo ad avere livelli differenti di attivazione e dunque di impiego del concetto. Dal punto di vista cerebrale, se si tiene ferma l'idea che a ogni concetto corrisponde un certo numero di neuroni, va giustificato il modo in cui tale *pattern* può variare di contesto in contesto quando si passa da una prima interpretazione ad una seconda e, in seguito all'abbandono di quest'ultima, si ritorna alla prima. Occorre forse ipotizzare la necessità di una soglia minima di attivazione mantenuta fra le varie interpretazioni? Oppure è più sensato ritenere che ogni volta che si passa da un contesto all'altro ci sia una sorta di azzeramento dell'attivazione e il processo di strutturazione ricominci da capo? Nel primo caso va comunque spiegato il fatto che «sembra esserci un'influenza del tutto trascurabile di un livello di attivazione in un contesto sul livello di attivazione dello stesso concetto nell'altro contesto» (*ivi*, p. 83), ovvero, che un soggetto umano può usare lo stesso concetto in due contesti diversi nello stesso processo di pensiero senza evidenti o frequenti problemi di sovrapposizione e interferenza, come mostra un ragionamento del tutto plausibile del seguente tipo: "per aprire la portiera di *sinistra* della macchina devo prendere la chiave che si trova nella tasca *sinistra* del mio cappotto". Nel secondo, diventa difficile spiegare come mai sia così facile e immediato ritornare ad una prima interpretazione una volta abbandonata quella successiva, ovvero si presenta il problema di come mai è possibile il recupero quasi immediato di una data visione delle cose, senza passare attraverso le fasi standard di costruzione della rappresentazione, a meno che non si ricorra a un qualche tipo di memoria. E allora la domanda diventa: quale tipo di memoria?

L'implementazione al computer offre una scappatoia a questo tipo di problema, cioè l'introduzione di un meccanismo di *stack* (pila) che conservi i "vecchi" *pattern* di attivazione, i quali sarebbero così in grado di entrare e uscire dalla visione corrente del programma attraverso processi di *push* e *pop*. Tuttavia, questo espediente sembra poco plausibile dal punto di vista psicologico a causa della nettezza (*clean*) dei simboli conservati in uno *stack* informatico (*ivi*, p. 88) e della potenziale ricorsività infinita di questo processo, fatte salve le risorse effettive del calcolatore, che garantisce un recupero totale dei dati immagazzinati. French, propende per un'altra soluzione tentando di conferire al programma la capacità di «ispezionare le strutture presenti nello Spazio di Lavoro e poi di *ricostruire* le vecchie attivazioni sulla base di ciò che ha osservato». Solo i concetti più astratti, infatti, devono essere azzerati, perché su di essi si costruisce un'interpretazione, mentre le strutture di più basso livello costruite (gruppi e corrispondenze)

[46] In realtà, un problema analogo è tipico dei sistemi connessionisti, quello del vincolo delle variabili rappresentate dai nodi di una rete neurale. Se ne discute, ad esempio, in Smolensky (1988).

rimangono in larga parte inalterate, essendo rappresentate nell'interfaccia dello Spazio di Lavoro da linee tratteggiate invece che continue. Questo non garantisce un processo di ri-attivazione perfetto, ma solo approssimativo, il quale appare, tuttavia, maggiormente plausibile dal punto di vista psicologico.

In conclusione, si può affermare che la nozione di interpretazione controfattuale che il programma gestisce è basata sulla possibilità del mantenimento di relazioni di raggruppamento e corrispondenza collegate a concetti astratti che sorreggono, univocamente o a *cluster*, interpretazioni alternative. Quella vittoriosa è quella che in ultima battuta fa sì che le corrispondenze e i gruppi perdenti siano cancellati dall'interfaccia dello Spazio di Lavoro e i concetti astratti corrispondenti siano portati a un quantitativo nullo di attivazione. TABLATOP è, dunque, in grado di simulare sia le attivazioni che rimangono in memoria, cioè quelle legate alle strutture direttamente collegate ai livelli più bassi del processo percettivo, sia l'azzeramento completo delle attivazioni che riguardano il livello concettuale astratto su cui è basata effettivamente l'interpretazione. Mentre le prime sono tendenzialmente tanto più fisse quanto maggiore è la concretezza degli elementi coinvolti (oggetti e relazioni di identità), le seconde vincolano la loro stabilità alla coerenza globale, in mancanza della quale passano repentinamente ad una fase di azzeramento per lasciare il posto a un altro gruppo di concetti che esprima l'interpretazione coerente del sistema.

La ricerca che ha prodotto il modello TABLETOP si è basata, come negli altri casi, su una serie di esperimenti su soggetti umani, cui sono stati sottoposti i quesiti di analogia. Questo tipo di metodologia è tipica dell'approccio simulativo dell'IA psicologistica fin dai tempi in cui Newell e Simon svilupparono il loro GPS (*General Problem Solver*)[47]. Tuttavia, va notata una distinzione rispetto alla metodologia basata sui resoconti introspettivi dei soggetti umani utilizzata dagli autori del GPS. Essa non è presente negli esperimenti compiuti, né potrebbe esserlo considerata la natura non conscia dei processi indagati, il livello *subcognitivo* appunto. Il confronto con le prestazioni, e non i resoconti, dei soggetti umani va verso una direzione di emancipazione dei residui introspezionistici tipici della prima fase dell'IA, che non vengono cancellati ma relegati alla fase intuitivo-creativa dei modelli simulativi che condividono questa impostazione. D'altra parte, la teorizzazione e l'implementazione al calcolatore di modelli simulativi dei processi creativi correrebbe il rischio di cadere in una sorta di circolarità se fondasse la sua metodologia effettiva sul fenomeno che intende spiegare, cioè quello dei processi intuitivi e creativi. Il ricorso al confronto con la prestazione a convalida della simulazione dei meccanismi cognitivi vuole essere un tentativo di limitare la circolarità esplicativa.

[47] Per una descrizione di questo programma e delle metodologie generali seguite nell'approccio dell'IA tradizionale di cui il GPS è il risultato più famoso si rimanda a Newell, Shaw, Simon (1959); Ernst, Newell (1969); Newell, Simon (1972).

In TABLETOP, tutto ciò assume una dimensione particolarmente evidente e proprio a causa del dominio scelto. Mentre nel caso dei problemi di analogia con le lettere o con le successioni numeriche, e anche nei domini geometrici che affronteremo in seguito, la soluzione da trovare scaturisce da un quesito che sembra, in maniera illusoria, implicare conoscenze specifiche (ad esempio, matematiche) nel caso del dominio di TABLETOP e dei suoi problemi di analogia appare evidente, più che nel caso degli altri modelli, che la conoscenza in gioco non è relativa a un elevato grado di *expertise*, bensì ai meccanismi, *e ai concetti*, implicati in senso astratto nella percezione definita di alto livello, un tipo di conoscenza che, in questa prospettiva, viene considerato un possesso generale sovra- e meta-contestuale del sistema. In tal senso vanno lette le seguenti parole di Hofstadter e French sullo scopo riconosciuto di questo progetto (e, in senso lato, anche di tutti gli altri) considerato in stretta relazione col suo dominio:

> Lo spazio dei *problemi* è, quindi, strettamente connesso con quello delle *pressioni mentali*, e in definitiva il progetto TABLETOP (in verità, ogni creazione di analogie) riguarda queste pressioni e le loro interazioni.
>
> L'obiettivo tangibile del progetto è quello di costruire un programma per fare «analogie da puntamento» entro questo piccolo dominio e in un maniera psicologicamente realistica. (Hofstadter, French, 1995a, p. 351).

Si potrebbe ipotizzare, come fanno Hofstadter e French (*ivi*. p. 381), che a questo scopo sia adatto anche un programma che sfrutti algoritmi di ricerca basati sulla forza bruta utilizzando un processo che prenda sistematicamente in considerazione tutti gli oggetti dello spazio percettivo (il piano del tavolo), assegni un valore ad ognuno e prosegua in modo ricorsivo confrontando l'oggetto che ha ricevuto il punteggio più alto con tutti gli altri. Il fine sarebbe quello di cercare le somiglianze in base a un qualche criterio specifico che richiede «un meccanismo progettato per tale scopo specifico» (*ibidem*) procedendo a un'azione di controllo che passi in rassegna tutti gli oggetti con punteggio discendente fino al momento in cui il confronto di quello scelto con tutti altri non dia esito negativo. Tale ricerca porta a risultati soltanto in una piccola parte dei casi presi in esame, anche se deve essere applicata sempre per ottenere un qualche risultato. Inoltre, lo scoprire i gruppi e le corrispondenze utilizzando algoritmi di ricerca di questo tipo significa avere a disposizione una molteplicità di meccanismi specifici «che equivarrebbe, in un ambiente realistico, al suicidio computazionale, e per di più costituirebbe un'assurdità dal punto di vista psicologico» (*ivi*, p. 383). Ne consegue, secondo la visione degli autori, che TABLETOP fa uso di un'architettura più potente, robusta e adatta per intervenire in problemi di questo tipo, nei quali «si considera, in una specifica dimensione concettuale, la *somiglianza* e non l'*identità*» (*ivi*, p. 367). Questo perché, a suggello di quanto detto finora, e con un richiamo agli aspetti psicologici fondamentali implicati nel progetto:

> [...] vi è una profonda interazione mutua tra i processi che costruiscono le nuove strutture e i processi che concentrano l'attenzione su determinati concetti e su determinate zone. *È il fatto di possedere questo tipo di architettura, ragionevole dal punto di vista psicologico, che impedisce a TABLETOP di subire un'"esplosione" combinatoria, non la piccolezza del suo dominio*. (*ivi*, p. 383 [enfasi mia])

Tale architettura, infatti, permette al programma di sfruttare al massimo il ruolo delle *pressioni selettive* concettuali nella strutturazione della situazione e nella costruzione della corrispondenza analogica. Fra esse, oltre alle posizioni e le dimensioni degli oggetti (aspetto spaziale), la categoria di appartenenza (aspetto categoriale) e i raggruppamenti disposti a più livelli (aspetto contestuale), va anche considerata un altro tipo di relazione istituibile fra gli oggetti, quella basata sulle «associazioni funzionali comuni» (Hofstadter, French, 1995b, p. 406) degli oggetti del dominio – come, ad esempio, il fatto che tazzina e cucchiaino spesso vengano usati insieme – le quali costituiscono un altro dei possibili modi in cui il pensiero crea delle corrispondenze nello spazio percettivo[48].

Per quanto riguarda la conoscenza detenuta e impiegata dal programma si può concludere che sia dotato di una struttura concettuale, e dunque una rappresentazione della conoscenza, piuttosto complessa, che ricomprende, come vedremo in maniera trasversale, molte delle caratteristiche delle differenti teorie dei concetti sviluppate negli ultimi anni. La rete semantica di TABLETOP, con i suoi differenti tipi di associazione concettuale e con la sua potenzialità dinamica, favorisce molte forme diverse di slittamento, basate sulle diverse tipologie delle correlazioni possibili fra concetti, e questo avviene sia che si parli di oggetti sia di gruppi di oggetti. L'unico aspetto di cui non è dotata questo tipo di rete, che potremmo definire *a collegamenti variabili*, è, ancora una volta, l'impossibilità di aggiungere nuovi nodi concettuali, che, come si è visto in precedenza, doterebbe questi modelli di una capacità di *learning* certamente maggiore. Tuttavia, i processi percettivi, emergenti, da essa diretti in modo probabilistico, rendono TABLETOP un programma in grado di muoversi anche in un ambiente in cui sono presenti «l'*indeterminatezza* e l'*ambiguità* proprie della vita reale, dove le situazioni non si presentano squadrate e impacchettate, ma vanno ritagliate con fatica dallo sfondo, mediante agenti percettivi che possono ampiamente differire tra loro per il modo in cui operano» (*ivi*, pp. 416-417 [enfasi mia]), cioè un ambiente molto più simile a quello in cui si muovono gli esseri umani e che presenta problemi le cui soluzioni non sono univoche ma possono variare anche *ceteris paribus*, sulla base della *semi-casualità intrinseca* all'architettura del modello. Tutto questo sarebbe impossibile senza la presenza di una rappresentazione della conoscenza che già di per sé inglobi sia l'indeterminatezza che l'ambiguità nella rete semantica del programma. Ciò costituisce la più vistosa differenza con COPYCAT e forse il tratto che denota il punto di maggiore evoluzione, rispetto a questo programma, di TABLETOP. Come fanno notare

[48] Questo aspetto richiama le teorie cognitive sui concetti basate sull'azione. Si consideri la seguente affermazioni della Borghi: «La percezione è selettiva in quanto estrae l'informazione funzionale dall'azione» (Borghi, 2002, p. 218). Si ritornerà su questo tema in maniera più dettagliata e ampia nel prossimo capitolo.

Hofstadter e French, «il Platbeto di TABLETOP (il suo repertorio platonico) è un guazzabuglio di concetti connessi in modo vago e confuso. Nella vita reale vi sono molti campi il cui repertorio concettuale è pieno di simili arbitrarietà e disorganizzazione, e contrasta con l'insieme di concetti più idealizzato e terso di COPYCAT» (*ivi*, p. 417).

3.6 Frammenti di alfabeti e lettere

3.6.1 La sfida dello stile

Finora abbiamo visto diversi modelli FARG, che costituiscono, ciascuno con le proprie peculiarità, un tentativo di utilizzare una metodologia simulativa per testare e comprovare i principi che regolamentano una visione subcognitiva dei fenomeni mentali. La discussione è stata condotta fin qui prendendo come punto prospettico di riferimento quello dei domini scelti come campo d'azione per i modelli. In quest'ottica, si può riscontrare che l'utilizzo di domini come quello degli stili alfabetici denota una tendenza verso domini sempre più raffinati e complessi, che implicano di conseguenza la teorizzazione e l'implementazione di modelli cognitivi in grado di fronteggiare situazioni caratterizzate da un grado crescente di dettagli e sfumature.

È il caso, ad esempio, di LETTER SPIRIT. Il modello, e il programma che ne è derivato, è congegnato per muoversi nel mondo degli stili alfabetici. A prima vista, la scelta di questo dominio non sembra presentare particolari differenze con quelli dei modelli precedenti, come nel caso dell'universo delle lettere (alfabetiche) di COPYCAT. Al massimo, si può pensare che sia soltanto un passaggio di livello dal considerare le lettere prese come entità passibili di raggruppamento al considerarle come forme autonome, passibili di una scomposizione in sottoparti. Tuttavia, proprio questo passaggio alla dimensione della scomposizione (e composizione) delle lettere pone una serie di problemi non solo più profondi per quanto riguarda la natura della percezione, ma anche più sottili in merito al livello dei meccanismi percettivi analizzati. Metaforicamente, si può dire che con LETTER SPIRIT si compia una discesa *all'interno* delle lettere, la quale costituisce, allo stesso tempo, una parallela discesa all'interno dei meccanismi cognitivi, che regolano *la possibilità del passaggio di livello* fra domini diversi della dimensione percettiva. Tale possibilità consiste nella capacità della mente umana di oscillare fra livelli gerarchici, ponendo attenzione ora all'uno, ora all'altro, con un processo non sempre del tutto cosciente, o, meglio, cosciente in una misura determinata dalle pressioni del contesto. Si è già visto cosa Hofstadter abbia detto in merito a questo problema da un punto di vista intuitivo parlando di meccanismi di tal genere (assenti) nella vespa Sphex[49]. LETTER SPIRIT costituisce un tentativo approfondito di simulazione di tali meccanismi,

[49] Ci riferiamo alla possibilità di salto di livello che gli agenti umani (e forse in alcuni casi anche gli animali) mettono in campo nel momento in cui si trovano davanti a un ostacolo (fisico, e perciò percepito, o, in altri casi, soltanto epistemico) nel tentativo di aggirarlo.

non visti soltanto come una risorsa fra le molte disponibili per aggirare un punto morto dell'elaborazione, come nel caso dello *snag problem* di COPYCAT e METACAT, ma inserito in una sofisticata architettura computazionale in grado di operare sulla base di continui passaggi di livello attraverso un ciclo (teorico prima che informatico) di retroazione fra vari livelli percettivi, perseguendo lo scopo della loro coerenza reciproca.

In questo modo si può leggere l'impresa complessiva di LETTER SPIRIT, modello sviluppato a più riprese e che ha dato origine a un'architettura complessa basata sull'interazione di diversi algoritmi interagenti a un livello macroscopico, o modulare, piuttosto che a un semplice programma unitario. Lo scopo principale del progetto sotteso a LETTER SPIRIT, infatti, (Hofstadter, McGraw, 1995) è quello di creare un sistema in grado di percepire, categorizzando, istanze di lettere, ma allo stesso tempo di produrre lettere che possano costituire un alfabeto coerente dal punto di vista stilistico, ovvero in grado di cogliere lo "spirito" complessivo che collega le une con le altre. È evidente come la fusione di questi due aspetti presenti una serie di problemi di difficile soluzione nel momento in cui essi vengano considerati separatamente e considerati in un ordine diacronico costituito da componenti completamente separate. Il riconoscimento di una certa istanza di lettera come appartenente a una data categoria avviene, infatti, sulla base di alcune scelte in merito ai costituenti della sua forma, e, dunque, in maniera ascendente, verticale, dalle parti al tutto. Tuttavia, la direzione del processo è anche considerabile come dal tutto alle parti, con la sistemazione di queste in un insieme coerente che faccia sì che l'istanza sia riconosciuta appartenere a una data categoria. Un processo del tutto analogo avviene nel caso di uno stile alfabetico, se consideriamo questo come un tutto e le lettere come parti che devono recare con sé tracce e indizi consimili al fine di generare una visione unitaria dell'alfabeto. Le pressioni dall'alto a una visione coerente possono spingere a modificare i singoli tratti di una lettera in modo che ricada all'interno dello stile generale percepito, ma anche facendole perdere in una certa misura le caratteristiche che la rendono più "vicina" a un certo tipo di categorizzazione in quanto lettera. Detto in altri termini, i processi di inserimento in una categoria e in uno stile, definibili entrambi attraverso una serie di tratti concettuali, sono molto facilmente a rischio di contrasto, tanto che Hofstadter e McGraw considerano le due dimensioni come «categorie fra loro ortogonali».

In tal senso, il modello sviluppato deve prevedere la possibilità di un continuo spostamento di livello fra "lettera" e "spirito", cioè fra il contesto implicato dalla categoria di lettera e quello relativo allo stile alfabetico, fino al raggiungimento di un equilibrio. Il processo non può non essere, dunque, *dinamico*. L'architettura del modello è chiamata a implementare un andamento ciclico di aggiustamenti progressivi fra i due livelli, chiamato «ciclo centrale di retroazione della creatività» (*central feedback loop of creativity*, Hofstadter, McGraw, 1995, p. 481), il quale non va considerato una semplice implementazione informatica di una struttura algoritmica, ma una sorta di meccanismo auto-regolatore delle interazioni fra i contesti. Non sfugge il richiamo alla TCCL del capitolo precedente, cioè la teoria del ciclo centrale cognitivo, come motore dei processi di pensiero

e assunto di fondo dell'approccio subcognitivo. Ciò collima anche con lo scopo generale del progetto LETTER SPIRIT, ovvero

> [...] il tentativo di costruire un modello per calcolatore di aspetti centrali della creatività umana, basato sulla convinzione che questa sia il risultato automatico dell'esistenza di *concetti fluidi*, cioè di concetti sufficientemente flessibili e sensibili alle influenze del contesto. (*ivi*, p. 437)

Così come per i modelli precedenti, è la fluidità concettuale a permettere la creatività nel dominio considerato, alla quale vanno, però, aggiunte alcune caratteristiche specifiche dell'architettura che permettano l'implementazione di un meccanismo generale di *passaggio fra contesti* e non soltanto una serie di pacchetti di informazione memorizzata che entri in azione nel momento in cui il programma va in stallo. Lo scivolamento fra contesti è, in definitiva, il prodotto centrale di questa architettura, di cui ora vedremo gli aspetti principali.

In primo luogo, va affrontata ancora una volta la questione del dominio. Se come si è detto, il passaggio alla comprensione della forma delle lettere, in senso categoriale, costituisce un arricchimento rispetto ai modelli che facevano delle lettere un universo chiuso e ben definito (COPYCAT e METACAT), l'espediente che viene trovato per superare la complessità del mondo reale è la semplificazione della struttura attraverso l'uso di una griglia[50] (fig. 3.4) che schematizza le lettere in costituenti minimi indivisibili detti *quanti*. Se una delle motivazioni addotte è quella di semplificare il dominio in cui opera il modello, essa non è l'unica, né la più importante. Infatti, la griglia definisce un ambiente percettivo che elimina l'elemento continuo presente, ad esempio, nei tratti curvilinei. Tale elemento, infatti, viene considerato di competenza di una ricerca sulla «visione di livello basso, o intermedio, e non l'alto livello concettuale» (*ivi*, p. 450), che costituisce l'obiettivo simulativo di LETTER SPIRIT. Gli aspetti di basso livello della visione delle lettere vengono considerati *superficiali*[51], di contro a una maggiore profondità e astrattezza della conoscenza coinvolta nella manipolazione dei tratti dello stile.

[50] La griglia è un rettangolo di 21 punti (un lato di 3 e uno di 7) collegati da linee tratteggiate. Ogni linea è un "quanto". Considerando anche quelli diagonali i quanti in totale sono 56, che sono anche tutte le possibili posizioni assumibili dai costituenti atomici delle lettere.

[51] La critica implicita nell'utilizzo di questa terminologia è nei confronti dei programmi che utilizzano regole (simboliche) di trasformazione di costituenti stereotipati dei caratteri per creare stili alfabetici differenti solo dal punto di vista della figura e non del concetto di lettera, come nel caso di DAFFODIL (Nanard, Nanard, Gandara, Porte, 1989), che si limita ad aggiungere a schemi di lettere forniti dall'esterno decorazioni fornite anch'esse dal programmatore, secondo un requisito di coerenza ma senza alcuna conoscenza concettuale profonda dei concetti delle lettere e dei loro costituenti, mancando in tal modo la possibilità di qualsiasi riferimento, rintracciabile nel programma, a meccanismi cognitivi. Per una discussione si rimanda a Hofstadter, McGraw (1995, pp. 438-441).

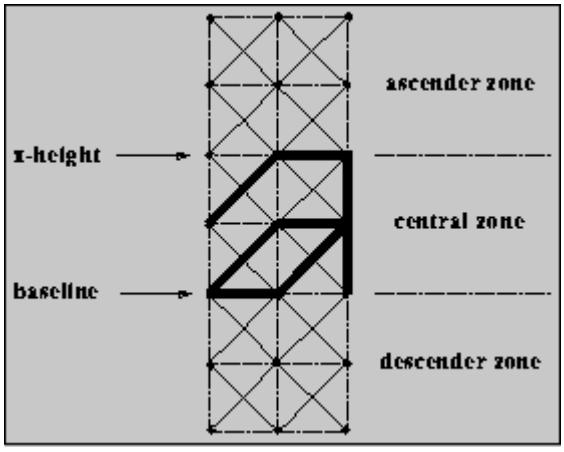

Fig. 3.4 - Rappresentazione grafica di un'istanza di lettera sulla griglia che definisce il dominio di LETTER SPIRIT (tratto da McGraw, 1995, p. 17)

Il presupposto teorico che sorregge questo tipo di ricerca nel mondo degli stili è che non esiste una forma precisa e definita che un'istanza deve assumere per corrispondere a un tipo. Esistono, ad esempio, infinti tipi di caratteri per esprimere la lettera "a" (*ivi*, p. 443). Il numero diventa finito se si considera la capacità limitata della griglia. Tuttavia, il punto saliente della questione è che si possono dare due istanze considerate della stessa categoria di lettera diverse sotto ogni aspetto (cioè, per ogni costituente) e riconoscibili solo sulla base di pressioni contestuali alfabetiche. Un comportamento interessante atteso dal modello, consisterà, dunque, nella capacità di categorizzare in maniera univoca queste istanze nella stessa misura in cui lo farebbe un agente umano. Questo appare, peraltro, un compito difficilmente assolvibile da un'architettura connessionista. Una rete multi-strato, infatti, in primo luogo è in grado di compiere in maniera molto efficiente compiti di categorizzazione, ma sembra trovarsi in difficoltà nell'applicazione di tratti stilisti espliciti, simbolicamente definiti. Secondariamente e in termini più generali, uno dei limiti delle reti neurali consiste proprio nell'impossibilità di mantenere distinti due contesti diversi e posti a un livello differente. Questo problema è collegato a quello delle interferenze catastrofiche cui va soggetta una rete addestrata nel momento in cui l'input tende a far divergere dalla media dei valori stabilizzati il pattern numerico che descrive la matrice dei pesi dei collegamenti della rete addestrata. Inoltre, se una soluzione opportuna potrebbe apparire quella di adottare un approccio modulare con una rete differente per ogni contesto rappresentato, si porrebbe pur sempre il problema della loro interazione[52], senza l'ausilio di un opportuno modulo di controllo simbolico esplicito attraverso cui

[52] Alcuni rilievi di questo tipo sono portati da Hofstadter e McGraw (1995, pp. 491-498) nella discussione che conducono in merito al modello connessionista GRIDFONT (Grebert, Stork, Keesing, Mims, 1991, 1992), costituito da una rete a tre strati *feedforward* e con *backpropagation* per l'apprendimento. Una critica estesa all'approccio connessionista al riconoscimento dei caratteri è presente in McGraw (1995, cap. 6) e in McGraw, Rehling, Goldstone (1994a), dove un'ampia serie di confronti fra esperimenti compiuti su soggetti umani e modelli al calcolatore, tra i quali alcuni appositamente approntati per questo tipo di sperimentazione (DUMREC, NETREC, NETREC+. Cfr. McGraw e Drasin, 1993) ha come conclusione la constatazione che la debolezza principale dei modelli connessionisti risiede nella mancanza di un apparato rappresentazionale simbolico di alto livello e flessibile, in grado di produrre una

far interagire *pattern* di dati strutturalmente omogenei, ma contestualmente dissimili. Un modulo di questo tipo presenterebbe, nondimeno, tutta una serie di problemi da risolvere e relativi alla sua plausibilità dal punto di vista psicologico. Infatti, se gli output dei due moduli connessionisti sono traducibili in forma simbolica, si può dire altrettanto del contesto che la rete rappresenta?

Sulla scorta di queste critiche si comprende ancor più la scelta di un dominio quantificabile in cui, tuttavia, non è richiesta l'uguaglianza della forma fra istanze di concetti di lettere. Ogni carattere, pertanto, viene riconosciuto, e prodotto, in base alle parti di cui è costituito, senza che ci sia nessun vincolo assoluto ad avere un certo tratto piuttosto che un altro, ma dipendendo questo fatto in ultima analisi dall'inserimento nel contesto dell'alfabeto di riferimento. Perciò, tanto per fare un esempio, la lettera "t" sarà soggetta a una descrizione che prevede la presenza di una barra verticale e di un trattino superiore che la taglia. Tuttavia, a fronte di pressioni stilistiche molto forti, il trattino può essere eliminato, nell'eventualità di un alfabeto il cui stile è caratterizzato principalmente dal non avere trattini orizzontali.

3.6.2 Un modello per il riconoscimento categoriale

Per costruire un'architettura in grado di compiere questo tipo di operazioni le lettere sono pensate in termini di ruoli e proprio allo sviluppo del modulo in grado di compiere questo tipo di analisi è dedicato gran parte del lavoro svolto nella realizzazione della prima implementazione di LETTER SPIRIT (McGraw, 1992, 1995; McGraw, Hofstadter, 1993, 1993b, 2002; Hofstadter, McGraw, 1995). Secondo questa impostazione, le lettere vengono considerate in forza di una quadruplice dimensione concettuale: il *concetto* di lettera, che esprime il concetto astratto «privo di forma» (Hofstadter, McGraw, 1995, p 442), che esprime la categoria cui si riconduce un carattere; la *concettualizzazione* della lettera, ovvero la sua scomposizione in termini di ruoli, i tratti descrizionali espliciti ("arco aperto", "barra verticale a destra", "trattino", e così via); il *progetto* di lettera, che è il modo in cui i ruoli sono realizzati e assumono una forma specifica sul foglio o nella griglia; il *carattere*, cioè la «forma grafica effettiva disegnata sulla carta, che realizza un certo progetto di lettera, quindi anche una concettualizzazione particolare e, in definitiva, un concetto di lettera» (*ivi*, p. 444). Questa suddivisione si riflette ai vari livelli in cui viene considerata la lettera: come intero, come insieme di ruoli e *r-ruoli*, cioè relazioni fra ruoli (punti di contatto o di intersezione, o anche estremi di ruoli), e come composizione di *parti grafiche* effettive.

In termini generali, si può dire che LETTER SPIRIT affronta il seguente processo di elaborazione. La prima fase è quella del riconoscimento dei caratteri che vengono dati in input attraverso un'analisi dettagliata, al fine sia di categorizzarli, sia di estrarne le caratteristiche stilistiche. In seguito, procede a disegnare le lettere mancanti per creare un alfabeto completo,

---

rappresentazione astratta dei caratteri percepiti. Tale critica viene estesa anche a modelli di riconoscimento basati sulla pura forza bruta e, per tale ragione, troppo rigidi.

cercando il più possibile di restare fedele a una visione stilistica coerente. L'analisi in termini di ruoli appartiene, appunto, alla prima fase dell'elaborazione. Essa consiste in una serie di processi interconnessi *bottom up* e *top down*. L'elaborazione comincia, come al solito, attraverso la scansione *bottom up* degli elementi percepibili, cioè i quanti sulla griglia. Questi vengono connessi in parti, strutture i cui aspetti salienti sono gli estremi e i punti di connessione con altre parti. A questo punto intervengono i processi *top down*, ovvero l'attivazione di ruoli e r-ruoli (relazioni fra ruoli), rappresentati in forma concettuale nella rete semantica, i quali tentano di adattare le parti alla loro struttura, grazie a una serie di specifiche, anch'esse in dotazione alla memoria del programma, che definiscono le *violazioni di norma* in base a cui un ruolo può variare per adattarsi al materiale percepito, i quanti strutturati in parti. Infine, l'attivazione dei ruoli influenza altri concetti nella rete semantica, gli *interi*, cioè i concetti di lettera che sono definiti, appunto, da una lista di ruoli e r-ruoli. Nel momento in cui l'elaborazione giunge all'attivazione di un concetto di lettera, che ricomprenda sotto di sé attraverso gli opportuni aggiustamenti dei ruoli intesi come regole (norme) standard tutto il materiale percepito presente nella griglia, la categorizzazione si può considerare compiuta e il riconoscimento della lettera terminato.

In LETTER SPIRIT sono presenti sia i meccanismi già visti negli altri modelli – le microprocedure esplorative e produttive, la scansione parallela a schiera che implementa processi probabilistici di ricerca e costruzione della rappresentazione, la temperatura come variabile di auto-regolazione della quantità di probabilismo dell'elaborazione – sia una serie di apparati peculiari del programma. Questi costituiscono un arricchimento *corrispondente* alla maggiore complessità dei processi percettivi simulati. L'aspetto più interessante da questo punto di vista consiste nella presenza di quattro diversi tipi di memoria implementati, che catturano la quadruplice dimensione concettuale attraverso cui è rappresentata una lettera nella conoscenza detenuta dal programma. Si hanno pertanto:

- una *Memoria Concettuale*, che contiene la conoscenza permanente del programma relativa al dominio in cui opera. Fra le altre cose vi sono i ruoli, «rappresentati come collezioni di norme, [...] che definiscono i limiti accettabili della caratteristiche fisiche molto semplici associate con le parti (altezza, curvatura[53], ecc.)» (McGraw, 1995, p. 148). Tali norme sono implementate nella rete in modo che quelle più tipiche siano più vicine al nucleo del concetto (cioè ricevano più attivazione) e viceversa. Nella rete sono rappresentati anche gli interi, come insiemi di ruoli e relazioni tra essi. Gli interi sono le concettualizzazioni delle lettere, implementate come descrizioni *simboliche esplicite prototipiche*. La rete è in grado di rappresentare gli aloni concettuali, cioè le relazioni di prossimità concettuale fra i vari

[53] Seppure le linee curve sono state espulse nella costruzione del dominio operativo di LETTER SPIRIT, vengono considerate parti curve quelle formate da una linea dotata di piegature a 45°. Si pensi, ad esempio, alla forma ad arco che può essere raffigurata da tre segmenti consequenziali (una diagonale a sinistra, una orizzontale centrale, una diagonale a destra) anche senza bisogno di un andamento curvilineo continuo fra i segmenti.

nodi che rappresentano in maniera simbolica i concetti (ruoli o interi che siano). Una particolare di questo tipo di rete è che essa è sì dinamica, nel senso già definito di propagazione dell'attivazione concettuale, ma non è dotata della facoltà di modificare le lunghezze delle proprie connessioni, la quale era una caratteristica, ad esempio, nella rete di TABLETOP;

- un *Centro Visivo*, che corrisponde allo Spazio di Lavoro dei precedenti modelli e in cui vengono create le strutture percettive;
- un *Blocco degli Schizzi*[54] (*Scratchpad*), sul quale vengono disegnati i caratteri, dall'abbozzo alla stesura completa, attraverso tutte le modificazioni compiute dal programma;
- un *Centro Tematico*, in cui si raccolgono tutte le idee che il programma ritiene definiscano lo stile dell'alfabeto, le "proprietà stilistiche", le quali saranno utilizzate nella produzione delle lettere mancanti rispetto a quelle input. Tali informazioni sono conservate esplicitamente sotto forma di *temi* prodotto nel momento in cui viene individuata una serie di tratti ricorrenti nei caratteri iniziali. Il Centro Tematico ricorda una struttura con una funzione per qualche verso analoga in METACAT, lo Spazio dei Temi, in cui si raccoglieva l'informazione al meta-livello concettuale rispetto a quello delle rete semantica.

In Hofstadter e McGraw (1995, pp. 467-468) vengono correlati questi tipi di memoria «con vari tipi di memoria – umana o computazionale – più familiari». Se il Blocco degli Schizzi e la Memoria Concettuale sono facilmente riconducibili, rispettivamente a un supporto esterno e a una memoria semantica permanente (una MLT), degna di nota è l'interpretazione che viene data del Centro Visivo e del Centro Tematico. Il primo è visto come «uno *spazio di lavoro subcognitivo* (cioè come una memoria di lavoro a brevissimo termine, e ad accesso assai rapido, come la memoria cache di un calcolatore), in cui processi percettivi paralleli agendo collettivamente e per lo più al di sotto della soglia di consapevolezza del sistema, stabiliscono una rapida classificazione superficiale di una forma e ne rendono accessibile dal punto di vista cognitivo la designazione finale di categoria». Il secondo «può essere pensato come una *memoria di lavoro cognitiva* (cioè una memoria di lavoro di tipo molto più conscio rispetto alla precedente, in cui si immagazzinano, si paragonano e si modificano le astrazioni derivate da percezioni più concrete e primarie». Questa caratterizzazione rende, dunque, esplicito, a differenza dei modelli precedenti, il ruolo parimenti necessario di processi subcognitivi e cognitivi. Non solo gli uni appaiono dipendere dagli altri in maniera reciproca, ma la loro interconnessione assume il ruolo di elemento necessario all'articolazione di processi percettivi sottili come quelli implicati dal dominio in oggetto, in cui lo spostamento continuo da un livello all'altro, entrambi rappresentati esplicitamente nel sistema, garantisce un risultato esteticamente raffinato e concettualmente ricco in termini parimenti di comprensione e produzione degli oggetti del dominio. A conferma di questo, può essere considerata

[54] Nella traduzione di Hofstadter e Graw (1995) viene reso in italiano con "scartafaccio".

un'affermazione incidentale fatta nella corso della presentazione generale del progetto LETTER SPIRIT, affermazione che contraddistingue l'intera impostazione di ricerca subcognitiva:

> È essenziale ricordare che le persone hanno in mente un insieme di *idee*, non un'*immagine*. (Hofstadter, McGraw, 1995, p. 444)

Al di là degli echi wittgensteiniani (in particolare in riferimento al *Tractatus*) sul complesso e ampiamente discusso rapporto fra raffigurazione e rappresentazione di un fatto, le parole riportate sembrano rimarcare una presa di posizione nel lungo dibattito fra visione immaginista e proposizionalista dei contenuti mentali, scaturito all'interno delle scienze cognitive già negli anni settanta del secolo scorso[55]. Esse sembrano potersi considerare un rifiuto della posizione immaginista, pur non essendo allo stesso tempo neppure una netta adesione al proposizionalismo dei contenuti mentali. Tuttavia, la struttura del centro tematico, nel quale sono rappresentate in maniera simbolica esplicita le proprietà stilistiche, potrebbe essere un indizio verso questa direzione interpretativa.

Tali questioni sono forse meglio inquadrabili alla luce dei principi della metodologia simulativa impiegata. Si considerino i criteri di selezione dei contenuti del centro tematico (Hofstadter, McGraw, pp. 456-457). Come nel caso dello spazio dei temi si tratta, infatti, di ricavare anche in questo caso opportune collezioni concettuali che descrivano le qualità stilistiche sulla base delle quali operare la costruzione delle lettere mancanti dell'alfabeto da completare. L'euristica alla base dell'estrazione delle proprietà stilistiche prevede l'estrazione di temi che riguardano, innanzitutto, la caratterizzazione dei ruoli. Infatti, essendo l'utilizzo dei ruoli, già definiti come l'aspetto normativo della conoscenza in gioco nel processo produttivo, soggetto alle varianti applicative, le così dette "violazioni di norma", proprio queste possono diventare un tema specifico, nel senso che una loro ripetizione in più caratteri dell'input può costituire una spinta alla loro replicazione nei nuovi caratteri. Altri due aspetti stilistici importanti sono i "motivi", forme geometriche ripetute più volte di carattere in carattere anche in misura parziale, e le "regole astratte", vincoli su specifici aspetti generalmente di basso livello e non relativi a forme particolari e "complesse" come i motivi. Come esempio si può considerare il seguente: "utilizza solo quanti verticali". Ma, a livello di teoria del modello, non vengono posti limiti espliciti alla formazione di regole possibili. Dipendono causalmente, così come le regole di trasformazione delle stringhe in COPYCAT, dai concetti attivati nel corso dell'elaborazione.

Appare chiaro, dunque, come tutti questi contenuti conoscitivi sono regole astratte specificabili in termini simbolici che riflettono il modo in cui un agente umano opera *scelte consapevoli* nell'eseguire la creazione di uno stile alfabetico (o grigliabetico). LETTER SPIRIT deve essere in

[55] Per un resoconto si rimanda a Luccio (1998). Per un'esposizione dettagliata e corredata di prove sperimentali delle due posizioni dal punto di vista immaginista si rimanda a Kosslyn (1980, 1983) e a Kosslyn (1994) e Denis, Mellet, Kosslyn (2004) per un'estensione del dibattito dal punto di vista del funzionamento cerebrale.

grado di utilizzare questa conoscenza nello stesso modo, nonché di produrla in maniera psicologicamente plausibile, attraverso un processo emergente, a superamento di soglia, di formazione concettuale, che in questo caso si situa a un livello molto astratto.

Tale compito, al pari di tutti gli altri nel modello, è lasciato all'operato delle microprocedure[56] che procedono all'esplorazione del materiale nello spazio percettivo, alla fusione del materiale atomico individuato in complessi strutturati, fino ad arrivare a compiere veri e propri processi di categorizzazione e di produzione di nuovi caratteri. Il funzionamento dell'apparato delle microprocedure rispecchia quello dei modelli già visti, con la scelta su base probabilistica gradualmente influenza da un quantitativo sempre maggiore di pressioni *top down*. Una novità degna di attenzione rispetto ai modelli precedenti viene introdotta nel modo di descrivere l'architettura.

Come abbiamo più volte affermato, il fatto che si possa dare una descrizione delle fasi dell'elaborazione del programma in termini di processi simbolici astratti ed espliciti non toglie nulla al fatto che essi siano svolti in realtà dall'azione dei micro-agenti o lavoratori. La descrizione di alto livello è presente nell'occhio dell'osservatore esterno al programma, o, al limite, nel caso dei modelli che sviluppano meccanismi auto-osservativi, nei moduli architetturali appositamente dedicati alla traduzione in termini simbolici espliciti delle macro-azioni compiute. Infatti, anche in LETTER SPIRIT sono presenti sia una conoscenza permanente (la rete semantica) sia una conoscenza emergente prodotta dai microprocessi elaborativi (ad esempio, i temi o le strutture create nello spazio percettivo). Non esiste una descrizione esplicita delle fasi di livello più alto compiute dal programma, che ha, d'altra parte, un andamento articolato, descrivibile in termini di: esplorazione, categorizzazione, estrazione di aspetti stilistici, applicazione di tali aspetti, creazione di caratteri, valutazione della loro coerenza ed eventuale correzione della loro forma.

Le fasi generali dell'elaborazione vengono ascritte come azioni specifiche di quattro meta-agenti, cioè agenti di alto livello (Hofstadter e McGraw, 1995, p. 474), che sono:

1) l'«*Esaminatore*» (*Examinator*), autore dei processi che avvengono nel centro visivo e preposto ai compiti di riconoscimento e categorizzazione di un carattere;
2) l'«*Astrattore*» (*Abstractor*), che rileva le proprietà dello stile e giudica, in termini di qualità stilistiche, la coerenza dei caratteri prodotti con quelli dati in input o prodotti in precedenza ;
3) l'«*Immaginatore*» (*Imaginer*), che operando «solo al livello astratto dei *ruoli*», predispone una concettualizzazione adeguata per un progetto di lettera in formazione. Il fatto che operi al livello dei ruoli sta a significare che «*non vi sono mai implicate forme*» (*ivi* p. 475), ma solo norme e violazioni di norme, concordemente con l'assunto già espresso in precedenza

[56] Nella prima implementazione di LETTER SPIRIT sono sedici ed è possibile considerarle secondo una complessità gerarchica crescente, da quelle dedite alla formazione di legami tra i quanti a quelle specifiche per l'adattamento degli interi (per mezzo dei ruoli e delle relazioni fra ruoli) al materiale percepito. Cfr, McGraw, 1995, p. 161).

che queste operazioni sono esclusivamente di natura mentale, e dunque appartenenti a una dimensione qualitativamente diversa da quella delle immagini;

4) il «*Disegnatore*» (*Drafter*), che produce effettivamente il carattere sulla griglia in base ai suggerimenti del progetto fatto dall'Immaginatore.

Questi macro-agenti portano un nome che esprime il loro aspetto funzionale. Ognuno, infatti, è pensato svolgere una delle quattro mansioni specifiche che costituiscono l'ossatura di LETTER SPIRIT, rispettivamente: l'attività percettiva concreta, l'attività percettiva astratta, l'attività concettuale di alto livello e l'attività di livello intermedio, che si situa, cioè, fra il percettivo e il concettuale. Tuttavia, Hofstadter e McGraw non tralasciano di fare una precisazione fondamentale:

> Spesso è utile parlare di queste attività emergenti come se fossero espletate da quattro moduli espliciti e del tutto separati che nell'insieme abbraccino il programma intero[...]. Li battezzeremo così: l'*Immaginatore*, il *Disegnatore*, l'*Esaminatore*, l'*Astrattore* [...]. Si ricordi, però, che *questi moduli sono solo finzioni utili nella descrizione del programma*, dato che ciascuno è un semplice sottoprodotto delle azioni di molti codicelli e che le rispettive attività sono tanto intrecciate da non potere essere districate e isolate in modo netto. (*ivi*, p. 474 [enfasi mia])

Il passo è di notevole importanza perché evidenza una differenza non molto rimarcata nella descrizione dei modelli precedenti, quella fra modello e programma e quella fra modello computazionale e architettura. Tale distinzione non è nuova nell'IA. Si pensi ad esempio alla distinzione classica in tre livelli proposta da Marr per la descrizione esatta di ogni progetto simulativo: il livello della teoria computazionale, quello algoritmico, quello dell'implementazione (Marr, 1982). Nei modelli subcognitivi descritti in precedenza una tale distinzione è assente. Generalmente le componenti computazionali trovano una controparte algoritmica esplicita, come nel caso, ad esempio, delle differenti memorie, la MLT e MBT, che diventano quasi simmetricamente la Rete Concettuale permanente e lo Spazio di Lavoro. Tale corrispondenza è presente anche in LETTER SPIRIT per quanto riguarda i quattro tipi di memoria. Tuttavia, se consideriamo i quattro moduli appena visti, essi non hanno *elementi funzionalmente equivalenti* all'interno dell'algoritmo. Al contrario, la loro individuazione è, in termini descrittivi, soltanto emergente, e possibile solo *dall'esterno*. Perciò, mentre una distinzione fra modello cognitivo e programma sembra abbastanza ovvia, se con programma si intende il livello dell'implementazione al calcolatore, meno scontata appare la distinzione fra modello e architettura algoritmica, se si considera che alcuni aspetti dell'architettura rispecchiano moduli funzionali del modello e altri no. Come operare e su che basi, una distinzione? Ed anche, che lezione si può trarre da una distinzione di tal genere?

Per iniziare, si può constatare che i moduli descritti hanno tutti una natura operativa, cioè denotano azioni complesse, non solo la cui implementazione di presenta come problema specifico per lo scienziato cognitivo nelle vesti di programmatore, o per un programmatore con mansioni apposite inserito in un progetto di scienze cognitive, ma anche la cui descrizione algoritmica richiede una traduzione in termini diversi, consistendo in questo buona parte delle potenzialità esplicative che ha il modello. Si potrebbe pensare allora che tutto possa essere espresso preliminarmente in termini algoritmici, che, per larga parte sono intuitivi e non interessano questioni relative alla programmazione effettiva, perdendo così la facilità della trattazione in cambio, però, di una maggiore esattezza descrittiva del meccanismo cognitivo analizzato. In altri termini, si sarebbe tentati di conformare gli aspetti del modello a quelli dell'algoritmo, instaurando una relazione di corrispondenza biunivoca. Ma, è lecito chiedersi, al di là della convenienza è possibile un'operazione del genere?

In realtà, la risposta sembra essere negativa e proprio per le caratteristiche dell'approccio simulativo in discussione. In una prospettiva di analisi dei fenomeni mentali in termini subcognitivi, sembra arduo rifuggire dal ricorso a un qualche tipo di elaborazione emergente, cioè probabilistica, competitiva e micro-modulare dal punto di vista procedurale, e non perché non sia possibile ottenere prestazioni in qualche maniera simili utilizzando altri approcci, bensì, piuttosto, per non diminuire le *potenzialità esplicative* di questi modelli in relazione ai fenomeni di cui costituiscono il tentativo di simulazione e spiegazione. La distinzione fra modello e architettura, perciò, sembra piuttosto configurarsi come un vantaggio, e proprio nella misura in cui si sostituisce a quella fra teoria computazionale e algoritmo, come si vede nella distribuzione su più livelli gerarchici di funzioni tutte riconducibili al livello algoritmico tradizionale. Infatti, è proprio perché l'architettura rappresenta la scomposizione differenziata su più livelli di operazioni complesse, che quella diventa una spiegazione plausibile di queste. Con due conseguenze valide dal punto di vista esplicativo. Da una parte, viene favorita in tal modo la comprensione del fenomeno (attività cognitiva) attraverso la realizzazione in termini di (sotto-)azioni effettive, che non soggiacciono tuttavia ai vincoli di una rigida composizionalità deterministica, ma sono espletate secondo procedimenti stocastici e probabilistici. Dall'altra, si mantiene una linea di continuità fra i due, attraverso l'analisi nei termini della nozione di emergenza basata su elementi dotati di un opportuno riferimento significativo e non, al contrario, su *pattern* numerici strumentali scollegati dai livelli emergenti, aspetto che si può considerare una delle principali ragioni della debolezza esplicativa del connessionimo.

Per quanto riguarda la realizzazione effettiva del modello, McGraw (1995) presenta l'implementazione di uno soltanto dei quattro macro-agenti: l'Esaminatore o Modello dei Ruoli (*Role Model*), predisposto al riconoscimento di caratteri sulla griglia dati come input e per il quale è necessario lo sviluppo di soltanto due dei quattro tipi di memoria descritti: il Centro Visivo e la Memoria Concettuale. Il processo, già visto nelle sue linee principali, viene presentato come la progressiva integrazione di aspetti *sintattici* e aspetti *semantici*, che corrisponde al processo di

istanziazione di concetti di ruoli e di insiemi di ruoli e r-ruoli, le lettere "intere", con l'obiettivo della "lettura" del materiale percepito. Il procedimento segue, dunque, quello consueto di costruzione di una rappresentazione strutturata dello spazio percettivo sulla scorta dei concetti della rete semantica. I quanti sulla griglia vengono raggruppati ed etichettati. Questa operazione produce l'attivazione di ruoli specifici, con eventuali violazioni della loro caratterizzazione standard, i quali attivano a loro volta gli interi fino ad arrivare al riconoscimento effettivo, l'attivazione di un solo intero, cioè di un solo insieme di ruoli ed r-ruoli. La distinzione fra aspetti sintattici e aspetti semantici esprime la differenza fra operazioni indipendenti dal contesto, per le quali si rivendica un'origine evolutiva e, dunque, uno statuto *innatistico*, e operazioni contestuali: «le operazioni sintattiche producono *aggregazioni indipendenti dal contesto* che - è presumibile – si presenterebbero nel corso dell'attività di un qualunque sistema visivo evolutosi per via naturale»; di contro, i processi di secondo tipo producono «*parti semanticamente regolate*» (Hofstadter, McGraw, 1995, p. 477).

Fra le microprocedure due sono particolarmente interessanti. La prima di esse è l'"esploratore gestaltico", che provoca un'immediata risposta di riconoscimento saltando le molteplici fasi esplorative costruttive e che intende simulare la capacità umana di cogliere, senza la mediazione dell'analisi delle parti, il carattere dal punto di vista della sua interezza. Esso riduce drasticamente il tempo di elaborazione, ma accresce le possibilità di risposta erronea o forzata. In generale, infatti, l'esaminatore è autore di buone prestazioni, anche se la percentuale diminuisce in presenza di caratteri molto lontani da una caratterizzazione tipica. Questo accade anche per la mancata implementazione in LETTER SPIRIT degli altri macro-agenti e dei moduli architetturali ad essi necessari. Ciò che non è presente è, perciò, l'apporto all'elaborazione delle pressioni contestuali dello stile. Inoltre, come per gli altri modelli una certa quantità di errori, imprecisioni e forzature sono connaturate al tipo di strategia di ricerca probabilistica impiegata. Una seconda microprocedura degna di attenzione è il "cercatore di sussunzione di ruoli", che, con un richiamo terminologico kantiano, indica un procedimento di inclusione di ruoli all'interno di un ruolo più grande, sovra-ordinato, espletando così una funzione di creazioni di gerarchie concettuali non pre-programmate nella rete.

Tralasciando ulteriori dettagli tecnici è interessante notare come il processo di integrazione fra aspetti semantici e sintattici viene qualificato come il superamento della contrapposizione fra parti (aspetto percettivo) e ruoli (aspetto concettuale). Tale prospettiva non è scollegata da precise teorie formulate nell'ambito della psicologia della percezione, di cui si può dire ne costituisce quantomeno la controparte in termini di esperimento simulativo. La definizioni delle strutture formate a partire da elementi percepibili quali i "quanti" richiama, ad esempio, le ricerche in percettologia compiute da Palmer (1977) sulle unità minime strutturali ipotizzate per spiegare la percezione delle forme e delle similarità tra forme. Tali unità minime avrebbero una funzione determinante nell'attuazione di processi percettivi gestaltici in forma totalmente *bottom up*. La differenza con questa teoria consiste

nell'introduzione di pressioni *top down*, che ricorda, d'altra parte e per certi versi, la teoria del "riconoscimento per componenti" di Biederman (1987).

La scelta di McGraw di impostare il processo di riconoscimento su unità strutturabili, le parti (*parts*), è supportata dal confronto delle prestazioni di diversi modelli della percezione delle lettere (McGraw 1995, p. 240-291). Ciò che risulta interessante ai nostri fini è la distinzione operata da McGraw fra approcci piatti o a-gerarchici (*flat*) e approcci strutturati (*ivi*, p. 241 e sgg.). I primi non prevedono strutture intermedie fra la categorizzazione e l'apprensione di caratteristiche. I secondi, al contrario, procedono all'individuazione di strutture intermedie, le parti appunto, che vengono concatenate nell'intero attraverso un processo di analisi categoriale che porta al riconoscimento. Tipici del primo approccio sono, abbastanza intuitivamente i modelli connessionisti. Nel secondo ricadono, fra gli altri, gli approcci di Palmer, Biderman e quello della Treisman (Treisman e Gelade, 1980) basato sulle caratteristiche (*feature*). Il Modello dei Ruoli, come nota McGraw arricchisce il processo di riconoscimento inserendo anche relazioni fra ruoli (r-ruoli) che costituiscono un livello aggiuntivo di mediazione verso la categorizzazione. Al di là dei risultati sperimentali su soggetti umani, che pure sembrano confermare gli approcci strutturati, è interessante notare come McGraw attribuisce alla sua impostazione, fra le altre cose, la possibilità di mettere in atto dall'alto, attraverso l'operazione di integrazione fra parti e ruoli e r-ruoli, il processo analogico che «guida lo sviluppo della produzione del carattere» (*ivi* p. 245). In definitiva, la presa di distanza da approcci *flat* sulla base della possibilità di avvalersi di strutture percettive simboliche ai fini del processo di categorizzazione e, in senso lato, di produzione di analogie, può essere considerata anche una netta presa di posizione, sulla scorta dei risultati prodotti da LETTER SPIRIT, contro gli approcci connessionisti non adatti a modellare processi di costruzione di analogie che gli agenti umani sono ritenuti compiere sulla base allo stesso tempo di elementi strutturali percettivi e di un bagaglio concettuale codificato simbolicamente.

In generale, l'approccio attraverso ruoli si può considerare un approccio eminentemente funzionale che non rinuncia, però, all'integrazione con meccanismi percettivi "aperti", cioè in grado di fronteggiare una certa indeterminatezza dell'input. Si può dire che già TABLETOP condivideva questo tipo di impostazione, visto che l'elemento principale per la valutazione della salienza di un oggetto era il ruolo che esso giocava all'interno di un gruppo costruito. Tale aspetto era superiore, nel processo di assegnazione di valore e dal punto di vista del compito analogico, anche alle relazioni di prossimità concettuale e vicinanza spaziale, che pure costituivano uno degli obiettivi della valenza simulativa di TABLETOP. In LETTER SPIRIT, la maggiore articolazione del dominio e la caratteristica di essere un sistema specifico per simulare la capacità di "attraversamento di livelli", le cose sono rivoltate e considerate da un punto di vista opposto. Mentre in TABLETOP uno dei passaggi principali del programma era quello di individuare che ruolo rivestiva un certo oggetto dello spazio percettivo, in LETTER SPIRIT l'operazione che viene compiuta è quella di adattare il materiale percepito a ruoli predeterminati e passibili di variazioni. In

tal modo, viene riconosciuto un peso preminente all'aspetto concettuale nei processi percettivi anche di basso livello. Un tentativo di mediazione fra le due impostazioni potrebbe consistere in un sistema che costruisce i ruoli come strutture nello Spazio di Lavoro, così che essi siano il modo in cui il programma imposta la sua visione delle cose relativamente all'elaborazione corrente, senza utilizzare pacchetti di conoscenza pre-programmati, ma rimanendo comunque a un livello astratto, percettivamente svincolato, quale è richiesto in genere dai processi di mappatura analogica[57].

3.6.3 L'architettura complessa del processo creativo

Lo sviluppo delle restanti parti di LETTER SPIRIT è illustrato in Rehling (2001), il quale punta ancora di più l'attenzione sul fatto che il micro-dominio di azione del sistema è un *dominio visivo*. Ciò non significa che vengano introdotte variazioni sul dominio o nei moduli teorici di cui il sistema si compone. È, semmai, un'altra testimonianza dell'arricchimento progressivo dei domini in cui operano questi modelli, la cui naturale conseguenza sembra essere quella di implicare la tesi secondo la quale attraverso l'impiego di microprocedure è possibile simulare livelli più bassi (e più immediati) nello spettro del processo percettivo. Tali considerazioni rilanciano una serie di questioni, su cui ritorneremo nella parte conclusiva. Ne possiamo, però, fin da adesso individuare due principali:

1. a quale livello del sistema mente-cervello vanno fatte corrispondere le microprocedure?
2. se esistono primitive percettivi e relazionali cui il sistema mente-cervello è in grado di reagire anche in maniera immediata (la fase *bottom up* di ogni elaborazione) le microprocedure vanno considerate corrispondere a questo livello operativo del sistema, o si deve pensare che ad esse sia possibile ascrivere un'effettiva libertà di azione inter-livello, così come postulato nei modelli presi in esame?

Naturalmente le risposte a queste domande dipendono dai vincoli che si pongono al modello in merito alla distinzione fra le parti che si ritiene abbiano un'esatta corrispondenza nella teoria cognitiva esaminata e le parti che sono soltanto strumentali al funzionamento del sistema, un problema piuttosto spinoso che interessa l'applicazione generale delle metodologie simulative allo studio dei fenomeni cognitivi. In Rehling, secondo questa prospettiva, proprio la questione della natura cognitiva delle microprocedure rimane aperta:

> [Un codicello] è una routine relativamente corta che esegue alcune piccole operazioni, nessuna delle quali fa una gran parte del lavoro del programma. Questa è soltanto una prospettiva di tipo informatico sui

[57] Si veda Linhares (2005) per un tentativo in questa direzione applicato al dominio degli scacchi, nel quale sono ben coniugate proprietà spaziali e operative degli elementi in gioco (i pezzi).

> codicelli – essi vengono anche considerati corrispondenti *in senso significativo* (*meaningfully*) a eventi cognitivi di piccola scala, sebbene non è stato dimostrato che equivalgano a elementi del pensiero umano reale. (Rehling, 2001, p. 167 [enfasi mia])

Nella seconda implementazione di LETTER SPIRIT[58] (Rehling, 1997, 2001) sono presenti alcune importanti variazioni rispetto alla prima.

In primo luogo, nel riprendere il discorso in merito alle dinamiche concettuali coinvolte, Rehling pone l'accento su una considerazione *olistica* della relazione parte-tutto: «con le lettere, come con molte altre cose, l'intero è più grande della somma delle sue parti» (Rehling, 2001, p. 164). Perciò, il senso in cui una categoria di lettere va considerata intera è quello per cui una definizione delle sue parti è insufficiente a determinarla univocamente. Era per supplire a tale mancanza che vennero inseriti nella memoria concettuale i ruoli relazionali (i già visti r-ruoli), i quali, nella seconda implementazione, hanno, però, un'individuazione nella rete e una descrizione operativa. Sono, cioè, test che sperimentano le condizioni *sine qua non* delle relazioni presenti in un "intero", la lettera come insieme di ruoli, e necessarie alla sua attivazione. Essi sono di tre tipi: quelli che valutano i contatti e le intersezioni fra i ruoli; quelli che testano la non vuotezza di ogni ruolo presente nell'insieme; quelli che garantiscono che ogni quanto della griglia sia "coperto" (*covered*), ovvero che abbia una spiegazione nel contesto della categoria di lettera scelta.

È da notare che in LETTER SPIRIT 2 alla Memoria Concettuale (a lungo termine e permanente) si aggiunge la Rete Concettuale, un meccanismo che ne costituisce una sottoparte e che nell'implementazione di McGraw veniva lasciato implicito, cioè non separato dal resto della memoria permanente. La Rete Concettuale è pensata come una memoria a breve termine che registra in un spettro di attivazione compreso fra -100 e 100 la misura in cui ruoli e insiemi di ruoli sono stati selezionati come concetti atti a "coprire" il materiale percettivo in input, o ci si aspetta che lo siano in base all'informazione mandata dalle microprocedure esplorative. L'aspetto interessante risiede, da una parte, nel fatto che la rete è divisa in due livelli, quello dei ruoli e quello degli insiemi di ruoli, e che tale «divisione [..] è esplicita, poiché le regole che governano la quantità di attivazione diffusa differiscono nelle due direzioni» (*ivi*, p. 166); dall'altra perché essa è, di fatto, un rete connessionista senza strati nascosti in cui la conoscenza è rappresentata in maniera locale e non distribuita, la cui differenza con l'implementazione della Rete di Slittamento consiste nella mancanza di cambiamenti dinamici nel corso dell'elaborazione[59]. Tuttavia, questo è l'unico esempio di utilizzo di un meccanismo connessionista in questi modelli e, fatto, ancor più

[58] D'ora in avanti questo modello, pur essendo una diretta continuazione del primo verrà chiamato LETTER SPIRIT 2, per facilitarne la distinzione.

[59] Si potrebbe pensare che le variazioni di lunghezza dei legami della rete di slittamento corrisponda alla variazioni dei pesi su una rete connessionista. Tuttavia, mentre la prima caratteristica può essere interpretata modellare esplicitamente la relazione di prossimità concettuale (sia essa spaziale, funzionale o di sovra-ordinamento categoriale), non sembra sia possibile attribuire una simile interpretazione alla variazione dei pesi dei legami della rete connessionista. Si veda quanto detto in precedenza parlando di TABLETOP e ancora French (1995, p. 59)

significativo, proprio con il compito di modellare il passaggio fra livelli concettuali nelle parti alte dello spettro dell'attività percettivo-cognitiva. I nodi della Rete Concettuale hanno tutti un'interpretazione simbolica e l'utilizzo di una rete connessionista sembra ascrivibile esclusivamente all'intenzione di produrre un comportamento emergente e non un'elaborazione sub-simbolica. Almeno per LETTER SPIRIT 2, perciò, si può parlare di architettura ibrida, seppur in maniera contenuta e limitata soltanto a una sua specifica componente funzionale.

Dal punto di vista del modello, le Rete Concettuale permette di simulare un aspetto del riconoscimento assente in LETTER SPIRIT e che richiama la distinzione fra processi coscienti e processi che avvengono sotto la soglia dell'attenzione cosciente. Infatti, mentre questa distinzione era esplicitata in riferimento, rispettivamente, ai processi collegati all'uso del Centro Tematico e a quelli relativi all'impiego del Centro Visivo, seppure del primo non era stata data un'implementazione, in LETTER SPIRIT 2 essa appare nuovamente nella differenziazione dei due livelli della Rete Concettuale e nella loro rappresentazione attraverso una rete connessionista. Il fenomeno cognitivo che si vuole simulare consiste nel fatto che, mentre il riconoscimento di un carattere viene considerato in genere un evento cosciente, ciò che accade prima appartiene alla dimensione dei processi subcoscienti, che solo dopo aver superato un determinato valore di soglia producono il riconoscimento effettivo del carattere in esame. Tuttavia, nel momento in cui tale riconoscimento non avviene, si avvia l'analisi, questa volta a livello cosciente, delle parti del carattere in esame, al fine di esprimere un giudizio esplicito di appartenenza categoriale, anche attraverso un procedimento di prova ed errore. Le due direzioni dell'attivazione della rete simulano proprio questa dinamica. Il riconoscimento avviene immediatamente laddove un certo valore di soglia viene superato attraverso un'attivazione diffusa ed equilibrata di tutti i nodi che convergono ad un nodo-intero al livello superiore. Nel momento in cui un nodo-ruolo è molto più attivo degli altri, impedisce l'attivazione di un nodo-intero facendo concentrare l'attenzione del programma su di sé. Tuttavia, in mancanza di forte attivazione sia di un nodo-ruolo, sia di un gruppo di nodi ruolo fino al superamento della soglia di un nodo-intero, sono proprio questi, i nodi che stanno per le lettere, a inviare attivazione ai "sottoposti" nodi-ruoli corrispondenti alla ricerca di un loro possibile riempimento, così come un agente umano farebbe nel momento in cui cercasse di adattare il materiale percepito al basso livello ai frammenti della categoria di lettera che sta sperimentando come possibile categorizzazione[60].

In LETTER SPIRIT 2, come si diceva, vengono implementati anche gli altri macro-agenti o moduli previsti nel progetto iniziale. Ancora una volta viene ribadito che essi «differiscono in modalità che sono dovute alle differenze fondamentali tra i compiti che devono espletare» (*ivi*, p. 213). L'impostazione che viene data all'architettura è, però, completamente differente rispetto a quella del progetto iniziale. In LETTER SPIRIT 2 i moduli vengono implementati come programmi distinti (Rehling, 2001, p. 291), che condividono alcuni tipi di memoria (Memoria

[60] Per questi aspetti più tecnici si rimanda a Rehling, Hofstdater (1997).

Concettuale, Rete Concettuale, *Focus* Tematico[61]), ma hanno propri Spazi di Lavoro e specifiche microprocedure. Questa trasformazione mette in atto una vera e propria rivoluzione: trasporta elementi considerati esclusivamente a livello del modello in LETTER SPIRIT all'interno del sistema, conformando il sistema alla struttura delineata a livello della teoria computazionale e trasformando il primo LETTER SPIRIT in un'architettura complessa del tipo, ad esempio, di quella del sistema SOAR (Laird, Newell, Rosenbloom, 1987; Newell, 1990). Ancora una volta, questa può essere considerata una dimostrazione di quanto sia labile il confine che separa livello della teoria computazionale e livello cognitivo al punto da essere largamente violato.

Consideriamo cosa producono questi cambiamenti nello sviluppo del progetto. L'Astrattore viene qui definito Aggiudicatore (*Adjudicator*) e ha il compito di rilevare le proprietà stilistiche che definiscono lo stile alfabetico da produrre, ricavandole dai caratteri dati in input, inseriti in una struttura di memoria nuova chiamata Biblioteca (*Library*) come rappresentanti peculiari dello stile. Inoltre, questo modulo-programma è chiamato a equilibrare le caratteristiche nel *Focus* Tematico e a valutare in che misura un carattere rispecchia un certo stile. Deve assolvere, insomma, a compiti che appaiono tra loro contrastanti. L'Aggiudicatore agisce riempiendo il *Focus* Tematico di proprietà stilistiche che ricadono nelle tre categorie summenzionate: motivi, regole astratte e violazioni di norma – istituendo una gerarchia di frequenza fra le proprietà ritrovate. Essendo un modulo separato, e non più una parte del modello emergente nell'elaborazione del programma, le strutture algoritmiche che utilizza sono solo in parte le stesse degli altri moduli. Ad esempio, nella Memoria Concettuale sono rappresentati anche concetti di proprietà stilistiche, ma l'Aggiudicatore agisce in un proprio Spazio di Lavoro dove crea corrispondenze fra ruoli e proprietà stilistiche, le quali sono, perciò, inserite nel *Focus* Tematico solo dopo che l'Esaminatore ha compiuto il suo lavoro di riconoscimento.

Nell'implementazione di LETTER SPIRIT 2 l'Aggiudicatore mostra alcune rigidità e mancanze, come l'incapacità di proporre proprietà stilistiche "nuove", invece di registrare soltanto quelle inviduate. Inoltre, la sua registrazione è seriale e cumulativa, né è stata sviluppata una funzione di rivalutazione di caratteri già accettati. Le sue potenzialità, come nota anche l'autore del programma, non superano quelle dell'individuazione di un certo grado di coerenza visiva. Sulla scorta delle informazioni da lui prodotte agisce, tuttavia, il Disegnatore, che ingloba anche alcune delle funzioni dell'Immaginatore di LETTER SPIRIT, non implementato funzionalmente né architettonicamente neppure in LETTER SPIRIT 2. Le prestazioni del Disegnatore non sono molto elevate e rispecchiano i tratti di coerenza visuale rintracciati dall'Aggiudicatore, senza tuttavia procedere a processi di revisione.

Più interessante ai nostri fini è il fatto che esiste un quarto modulo, un programma di controllo di alto livello, chiamato Letter Spirit[62], che sovrintende all'attività dei tre moduli appena visti,

[61] Con questo nome viene chiamato il Centro Tematico, già descritto ma non implementato in LETTER SPIRIT.
[62] Useremo il carattere minuscolo per distinguere questo modulo dal sistema considerato nella sua interezza.

collegandone le rispettive attività in due fasi. La prima è quella di analisi e categorizzazione di un carattere in input. La seconda non è altro che il già descritto "ciclo centrale di retroazione della creatività" che ha inizio con l'attività del Disegnatore sulla base delle proprietà stilistiche selezionate nella prima fase. I suoi prodotti sono sottoposti all'attenzione dell'Esaminatore e dell'Aggiudicatore secondo un processo continuo di ri-osservazione e revisione, e inseriti nel Blocco degli Schizzi. Se la valutazione è positiva il carattere diventa la versione corrente della categoria corrispondente. Il ciclo viene ripetuto più volte per ogni lettera fino alla creazione di un alfabeto. Tuttavia, l'intero processo è soggetto a variabili contestuali e a un andamento non deterministico, per cui nonostante i reiterati tentativi «*il problema essenziale è che non è garantito*» (*ivi*, p. 316) un buon risultato. Questo, come si è visto in precedenza, è il prezzo da pagare in cambio della quantità di casualità che connota questo tipo di elaborazione. Tuttavia la casualità è alla base del processo creativo, che, a livello teorico, viene ricondotto a un ponderato equilibrio di una duplice euristica globale basata sulle nozioni di *evoluzione graduale* e *rivoluzione catastrofica* nella scelta di proprietà stilistiche per la costruzione dei caratteri (*ivi* p. 329).

La scelta del nome "Letter Spirit" in questa seconda versione del progetto, seppur non molto felice per la confusione che può ingenerare, mostra, d'altro canto, lo spostamento verso l'alto del livello della teoria computazionale, o, in altri termini, il fatto che non sempre esiste un solo livello al quale indagare un fenomeno cognitivo complesso[63]. Piuttosto esso, in questo caso, appare passibile di suddivisioni in più strati gerarchici, algoritmizzabili separatamente. Inoltre, poiché Letter Spirit è una sorta di meta-controllore dell'attività degli altri programmi e svolge le sue mansioni attraverso due fasi algoritmiche programmabili in maniera abbastanza semplice attraverso tecniche classiche di IA (in particolare attraverso algoritmi specifici di pianificazione), si può affermare che è un'ulteriore conferma che LETTER SPIRIT 2 può essere visto impiegare un approccio complessivamente ibrido. Le euristiche e le metodologie algoritmica impiegate dal programma variano a seconda del livello del fenomeno cognitivo osservato, il quale ricade ancora, tuttavia, all'interno dello spettro dei fenomeni definiti di "percezione di alto livello". Infine, appare chiaro che uno dei requisiti necessari all'impostazione ibrida risiede proprio nel modularismo del sistema che riflette la teoria cognitiva messa alla prova e sviluppa allo stesso tempo aspetti soltanto impliciti nei modelli precedenti. Ciò è vero almeno per quanto riguarda la creazione di livelli operativi simbolici "alti" che rispecchiano plausibilmente le capacità di pianificazione di macro-azioni che un soggetto umano mette in atto nell'affrontare un compito complesso.

Si noti che non alludiamo in questo caso alle capacità di auto-osservazione già viste in METACAT e, in prospettiva, in SEQSEE, le quali possono essere considerate risiedere allo stesso

[63] Sulla fusione fra i due livelli si considerino ancora ciò che Rehling afferma nell'elencare le tre prospettive in base a cui si propone di valutare il sistema. Di ogni sua parte vanno vagliate:

1. la plausibilità cognitiva;
2. la misura della qualità dell'output;
3. l'«efficienza dell'elaborazione del *modulo*/ *programma*» (Rehling, 2001, p. 338 [enfasi mia]).

livello delle altre attività. Infatti, in quel caso si trattava di *pattern* concettuali o concettualizzazioni di eventi, oggetto degli stessi meccanismi procedurali che operano indifferentemente su tutte le strutture di rappresentazione della conoscenza del programma. Nel caso di LETTER SPIRIT 2 sono soltanto *azioni*, cioè *aspetti operativi e non contenuti concettuali*, che consentono una suddivisione effettiva in livelli, suddivisione che non è colta da meccanismi di auto-osservazione, a meno che non si introduca un apposito dispositivo funzionale di, potremmo dire, *reificazione concettuale*. Tuttavia, ciò sembra andare oltre gli obiettivi di LETTER SPIRIT 2, il quale, in definitiva, *mette in atto azioni complesse senza la possibilità di meta-conoscere questo stesso atto*.

A uno sguardo complessivo, l'analisi di LETTER SPIRIT 2 mostra che l'obiettivo del progetto appare raggiunto solo in parte. Se, infatti, il sistema è dotato di una buona capacità di riconoscimento e categorizzazione relativamente al suo dominio, l'obiettivo più ampio di produrre un programma in grado di individuare e generalizzare stili alfabetici è conseguito in maniera lacunosa, al confronto delle prestazioni che può esibire un soggetto umano: «in generale, le persone hanno l'abilità di astrarre stili in modi che sono molto più flessibili dell'abilità esibita da LETTER SPIRIT [2] nel fare questo. Le persone possono continuamente oltrepassare le limitazioni percepite di un dominio per creare in maniera originale, mentre LETTER SPIRIT [2], per lo più, combina un insieme finito di *proprietà stilistiche primitive* in un insieme di stili che si mostra ampio, ma con limiti che appaiono distintamente ristretti a un osservatore umano» (*ivi*, pp. 351-352 [enfasi mia]). D'altra parte, tali limitazioni sembrano attribuibili al fatto di operare attraverso "proprietà stilistiche primitive" senza un'adeguata implementazione di ulteriori e più raffinati meccanismi di retroazione, considerata uno dei fattori determinanti della diffusione coerente della creatività, mentre non sembrano ricollegabili alla trasformazione del sistema in un'architettura modulare dotata di uno specifico modulo superiore per il controllo, del quale pure viene sottolineata la mancanza di elasticità (*ivi*, p. 370).

D'altra parte, questa caratteristica allontana il sistema dai principi tipici dell'implementazione dell'approccio subocognitivo (l'architettura FARG), che tende a livellare le azioni e a gerarchizzare la conoscenza che il programma detiene sia della situazione nell'ambiente percettivo sia delle sue azioni effettive nel corso della strutturazione dell'ambiente stesso. Tuttavia, resta aperta la domanda se sia possibile far collassare il livello superiore di controllo, rendendolo emergente, nelle operazioni dei moduli sottoposti, che condividono parziali strutture operative procedurali e di memoria, e mostrano anche una competenza sovrapposta nel "maneggiare" caratteri alfabetici[64] dal punto di vista della prestazione. Quest'ultimo aspetto in particolare indica che i vari moduli sono individualmente più efficienti, nel loro rispettivo compito, su stili alfabetici diversi. Ciò può suggerire l'idea che la scelta di un modulo di controllo superiore sia necessariamente implicato *a livello teorico* dal fenomeno simulato e non solo una scelta dovuta a un particolare tipo di

[64] Per una discussione sulle prestazioni incrociate dei moduli si rimanda a Rehling (1997). In Rehling (2001, p. 311) si trova una rappresentazione insiemistica che mostra i gradi di *performance* di ogni modulo e la loro sovrapposizione.

implementazione, e che, dunque, la ri-modellazione cui è soggetto il progetto sia dovuta a fattori strutturali legati all'obiettivo di sviluppare un compito cognitivo che opera su due livelli. In altri termini, l'ipotesi generale che si può avanzare da una ricognizione dei mutamenti nel passaggio da LETTER SPIRIT a LETTER SPIRIT 2 è che *simulare l'influenza di un doppio contesto* richieda necessariamente l'articolazione del modello in un'architettura cognitiva gerarchica che si riflette in una conseguente traduzione algoritmica dello stesso tipo. Ciò appare inevitabile ancora di più nel momento in cui tale doppio contesto ingenera pressioni divergenti, che il sistema deve porre in equilibrio.

In conclusione, è proprio per espletare il *suo*[65] compito analogico che LETTER SPIRIT 2 richiede un tale tipo di architettura. Ciò equivale ad affermare che esiste un legame molto stretto fra l'attività generale del compiere analogie, come attività cognitiva alta e basata su elementi simbolici discreti e strutturabili, e la presenza di un meccanismo ciclico di retroazione fra più livelli. Inoltre, tutto questo sembra essere consistente con la teoria definita del "ciclo centrale cognitivo" (TCCL), esposta nel precedente capitolo, che aspira a essere apparato esplicativo onnicomprensivo di tutti i fenomeni cognitivi di alto livello nello spettro delimitato a un estremo dalla categorizzazione e all'altro dalla creazione di mappature analogiche e di analogie in generale. Nella TCCL il ciclo è reso possibile dalla presenza di due tipi diversi di memoria, che, a questo punto, possono essere definiti in maniera più generale, come strutture di dati differenti e interconnesse, che variano dinamicamente e in modo diverso. Tali strutture sono vincolate verso l'"alto" da un certo grado, minimo, ma non inesistente, di *permanenza semantica concettuale* e verso il "basso" dalla stabilità dell'input scelto di volta in volta, stabilità che si riflette nelle modalità implementative perché conforme all'assunto di una capacità sintattica innata, immediata ed evolutasi naturalmente, procuratrice degli elementi percettivi di basso livello.

Due cose a questo proposito vanno, infine, menzionate. In primo luogo, Rehling individua le possibilità di auto-osservazione di LETTER SPIRIT 2 nella presenza di livelli intermedi di memoria piuttosto che nella constatazione che ogni modulo agisce sulla base dei risultati di un altro (*Rehling*, 2001, p. 367). I livelli intermedi, pur assolvendo funzioni diverse rispetto a quelli presenti in METACAT, sono strutture di dati che dotano il programma di quella complessità nella rappresentazione della conoscenza indispensabile per generare un ciclo di retroazione creativo e, dunque, per espletare il compito analogico di produrre uno stile alfabetico. La loro forza risiede infatti nella rapida variabilità che sono in grado di esibire:

> Lo stile dell'alfabeto [di griglia] in formazione è immagazzinato nel *Focus* Tematico e nella Biblioteca, mentre l'alfabeto stesso è immagazzinato nel Blocco degli Schizzi (che in termini di modellizzazione del comportamento umano, è probabilmente più giusto pensare come un pezzo di carta virtuale [cioè un

[65] L'attività di creare uno stile alfabetico viene considerata avere «la *forma* propria di un problema di analogia, sebbene mettere insieme così tanti elementi da ogni parte lo rende atipico»; esso è, dunque, da considerare «*come* un fare analogie» (Rehling, 2001, p. 359).

> supporto esterno] piuttosto che come una *rappresentazione mentale*. Questi sono tipi di memoria "leggi-scrivi" (a differenza della Memoria Concettuale che è immutabile) e possono perdurare immutati attraverso una scala di tempo di più di pochi secondi (a differenza dello Spazio di Lavoro). LETTER SPIRIT [2] modella effettivamente un compito "più grande" di quelli di COPYCAT e TABLETOP, e i livelli di memoria aggiuntivi riflettono questa complessità maggiore. (*ivi*, pp. 362-363 [enfasi mia])

In secondo luogo, per quanto riguarda i processi di percezione il quadro concettuale in cui si inserisce LETTER SPIRIT 2 riprende quello di LETTER SPIRIT, lo sviluppo del quale, relativamente all'implementazione del processo di riconoscimento, intendeva essere una dimostrazione della superiorità di rappresentazioni strutturate rispetto a quelle a-gerarchiche (*flat*), superiorità che veniva considerata valere, in prospettiva, anche per gli aspetti percettivi di alto livello. Rehling (2001, cap. 2) riprende questa prospettiva, ipotizzando, però, che l'attività percettiva sia svolta da un sistema più complesso, che incorpora i due tipi di meccanismi in modo complementare (*ivi*, p. 45). Le rappresentazioni strutturate interverrebbero in compiti percettivi in cui l'input è *topologicamente* simile alla categoria rappresentata nella mente. Le rappresentazioni a-gerarchiche sarebbero presenti nei fenomeni percettivi caratterizzati da un'estrema rapidità e in cui i confini, le cesure e i contatti nell'input percepito rivestono minore importanza nel processo di riconoscimento. Perciò, un lettera disegnata con tratti non continui verrebbe percepita da questo secondo meccanismo, in grado di riempire le relazioni mancanti.

A conti fatti, si tratta della ben conosciuta contrapposizione fra approccio analitico e gestaltico alla percezione, ognuno riservato a particolari situazioni presentate dall'ambiente. Il primo meccanismo supplisce, entrando in azione, alle inefficienze del secondo[66]. Ciò che è interessante, dal punto di vista della costruzione del modello, è il fatto che per denotare i due meccanismi venga usata una terminologia che richiama l'architettura proposta per il modello stesso (*ivi*, p. 51). Il primo meccanismo viene definito "algoritmico" (*algorithmic*), mentre il secondo "distribuito" (*distributed*). Tale distinzione richiama quella fra approccio simbolico e approccio connessionista all'IA. Tuttavia, l'autore chiarisce che non si tratta di un'esatta corrispondenza fra meccanismo algoritmico e processi seriali informatici, e fra meccanismo distribuito e connessionismo. Si vuole soltanto «suggerire che esistono certe somiglianze» (*ibidem*) e, dunque, questi termini vanno intesi al livello della teoria computazionale e non dell'implementazione. Al contempo, però, non sfugge il fatto che la gerarchia dell'intero apparato epistemico del sistema si dispiega su uno spettro tanto largo da contenere elementi di entrambi gli approcci, attraverso l'unificazione di un nucleo centrale dell'architettura fondato sulle specifiche idee del FARG.

[66] L'efficienza in termini di velocità del secondo meccanismo viene anche addotta a motivo della mancanza di ulteriori meccanismi auto-osservativi in LETTER SPIRIT 2, il cui impiego oltre una certa misura è ritenuto implausibile dal punto di vista psicologico. Lo sviluppo di capacità auto-osservative, infatti, ha come prezzo quello dell'elevata quantità di risorse computazionali (intermini di memoria e tempo) che la loro simulazione necessita (Rehling, 2001, p. 50), che rispecchia il dispendio di risorse mentali impiegate in compiti di questo tipo.

Il quadro complessivo che ne deriva sembra includere, dunque, una doppia dimensione (fig. 3.5): da un parte si hanno diverse strutture di memoria, che fondono componenti statiche e dinamiche a seconda del tipo di conoscenza che contengono o che sono predisposte a costruire nel corso dell'elaborazione. Ortogonalmente, e non parallelamente, a questa dimensione c'è l'insieme delle azioni possibili del programma dotate di un grado crescente di complessità, alcune delle quali *rappresentano funzionalmente*, cioè eseguono, azioni molto semplici e basilari, mentre altre *rappresentano funzionalmente* attività cognitive molto sofisticate. Così come dinamicità e permanenza delle strutture di memoria possono essere caratteristiche di ogni livello dell'elaborazione, simmetricamente senza meccanismi predisposti ad attuare la variabilità vincolata della conoscenza posseduta dal programma è impossibile attuare il duplice processo di riconoscimento categoriale e di costruzione analogica. L'intreccio delle due dimensioni è un ulteriore elemento a conferma della condivisione di un minimo comune denominatore da parte delle due attività "creative" di percezione di alto livello, che trova una esemplificazione in termini simulativi nel modello dell'architettura complessa di LETTER SPIRIT 2.

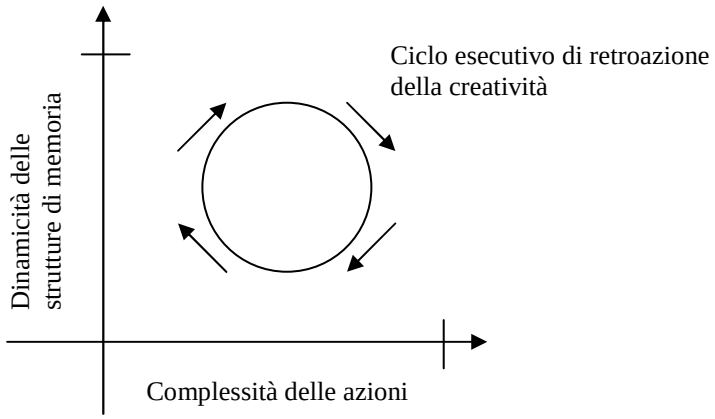

Fig. 3.5 - Le due dimensioni di LETTER SPIRIT 2

3.7 La geometria come problema limite dell'analogia

Un sistema di *analogy-making* recentemente sviluppato secondo l'approccio subcognitivo tipico del FARG è dedicato ai Problemi di Bongard. Come abbiamo visto in precedenza, i Problemi di Bongard sono costituiti da due serie di sei riquadri, ognuna contenente una certo numero di figure e in relazione di analogia fra loro secondo una caratteristica generale che differenzia i riquadri della prima serie da quelli della seconda. Ad esempio, ritornando al Problema di Bongard n. 71 della figura 3.1, la soluzione è data dall'individuazione del fatto che mentre in ogni riquadro di destra c'è

almeno una relazione di inclusione con al massimo un (ma andrebbe bene anche l'affermazione più forte "con esattamente un") livello di inclusione, in ogni riquadro di sinistra c'è almeno una relazione di inclusione con almeno due (ma andrebbe bene anche l'affermazione più forte "con esattamente due") livelli di inclusione. Tuttavia, in altri casi la soluzione dipende dalla forma delle figure o anche dalla presenza o mancanza di aree colorate piene e non colorate vuote[67].

Lo scopo di Bongard nel costruire i suoi problemi era quelli di testare le capacità umane nel compito di *pattern recognition*. Come abbiamo visto, proprio questo è il problema che Hofstadter (1979) proponeva come prototipico, e allo stesso tempo decisamente arduo, che l'IA doveva affrontare nella speranza di poter affermare di aver prodotto un programma "intelligente". Infatti, il riconoscimento di schemi (*pattern*) è strettamente collegato col tema della rappresentazione di una situazione percepita e con le questioni della categorizzazione, e di conseguenza connesso, attraverso il problema della conoscenza contestuale, con la questione della rappresentazione della conoscenza in un sistema umano o artificiale. Lo sviluppo di un modello computazionale traducibile in programma in grado di risolvere i problemi di Bongard fu, perciò, considerato da Hofstadter alla fine degli anni settanta del secolo scorso, una sorta di *experimentum crucis*, di spartiacque fra una visione dell'IA psicologistica troppo coinvolta con il simbolismo delle argomentazioni logico-algoritmiche e una visione che cominciava a occuparsi degli aspetti percettivo-cognitivi, da una parte in relazione al *come* della costruzione e della manipolazione della rappresentazione, e dall'altra con la dovuta attenzione alla questione non aggirabile del contesto epistemico in cui la rappresentazione viene operata. I Problemi di Bongard appaiono un ottimo banco di prova per tutti questi temi. Risolverli, trovare una possibile soluzione, vuol dire affrontare problemi di analogia e di riconoscimento interdipendenti e senza che tra i due compiti ci sia necessariamente una linea di separazione netta. Inoltre, il fatto di esprimere la soluzione in forma linguistica, se non implica il dover implementare un sistema per la produzione del linguaggio naturale, richiede tuttavia che si presti attenzione alla selezione dei concetti, in alcuni casi anche molto astratti, sulla base dei quali l'analogia viene contemporaneamente compiuta e spiegata. In quest'ottica, percezione delle raffigurazioni e analogia concettuale sono le due facce della stessa unica medaglia costituita dall'attività di riconoscimento di schemi (*pattern recognition*).

Dovrebbe essere ormai evidente la complessità del dominio costituito dai problemi di Bongard, da cui dipendono anche nel caso del nuovo modello proposto le particolari variazioni dell'architettura cognitiva ideata per risolverli: PHAEACO (Foundalis, 2006). Il punto centrale va ancora ricercato nel dominio. La risoluzione dei Problemi di Bongard implica al tempo stesso una serie di conoscenze relative alle figure geometriche, ma anche a relazioni spaziali e concettuali fra

[67] In Bongard (1970) sono esposti i cento problemi ideati dallo psicologo russo. In un manoscritto non pubblicato del 1977 (disponibile presso il Center for Research on Concepts and Cognition dell'*Indiana University*) sono raccolti altri 56 problemi ideati da Hofstadter. Per una consultazione molto più rapida di questi ed altri problemi proposti nel corso degli anni da differenti creatori (in tutto più di 250) si rimanda al seguente link: http://www.cogsci.indiana.edu/farg/harry/res/bps/bpidx.htm

esse, nonché la possibilità di vedere l'inesistente (spazi concavi, figure prodotte dal congiungimento ideale di punti, ecc.), gruppi di elementi non esplicitamente correlati. Le primitive percettive coinvolte sono, dunque, molto differenti: largo, piccolo, concavo, curvo, verticale, orizzontale, ma anche sotto, sopra, su, giù, uguale, diverso. In altri termini due sono le condizioni principali per ottenere una risposta a questi problemi:

- una conoscenza intuitiva della geometria, che possa essere facilmente collegabile con concetti astratti in modo da attuare relazioni di confronto (*matching*) a un qualche determinato livello;

- la possibilità di accedere agli elementi dell'input in maniera diretta e flessibile, per poter utilizzare l'informazione percepita, davvero cospicua, nel modo più fruttuoso possibile.

In particolare, la seconda condizione è legata al fatto che la discrepanza apparente tra ristrettezza del dominio impiegato e il mondo reale tende a diminuire nel caso dei Problemi di Bongard, a causa della complessità concettuale delle caratteristiche percepibili in gioco e delle strutture rappresentazionali coinvolte nei processi di *matching*. D'altra parte, tale discrepanza è stata definita "apparente", poiché, come abbiamo più volte affermato, nell'approccio subcognitivo non è la quantità di informazione presente nel micro-dominio a differenziarlo dal mondo reale, bensì piuttosto le capacità cognitive che devono essere messe in gioco per arrivare ad un adeguato svolgimento della prestazione che definisce il compito nel dominio. In relazione a questo criterio i Problemi di Bongard, in una ideale scala di complessità relativa all'insieme delle facoltà che mettono in gioco, sono il dominio più complesso visto finora. La sua complessità non deriva dal fatto di agire in dominio visivo molto ricco di informazioni, bensì in un universo i cui elementi sono soggetti a un tipo di "manipolazione" che implica una ricchezza percettiva (di oggetti e relazioni) anche ai livelli più bassi, come quello visivo. Così, mentre in COPYCAT gli elementi percettivi (le lettere) erano univoci e in LETTER SPIRIT (i caratteri), venivano vagliati, ma da un'angolatura che li vedeva sempre come istanze di un qualche tipo della sovra-categoria generale "lettera", in PHAEACO si arriva alla totale cecità pre-elaborazione in merito agli elementi del dominio, i cui unici vincoli, conosciuti dal programma, sono quelli di essere bidimensionali e racchiusi all'interno di un *frame* predefinito costituito da dodici riquadri suddivisi in gruppi di due. Per cui, se da una parte è vero che «lo scopo di PHAEACO *non* è quello di fornire un modello riuscito per l'automazione della percezione visiva o l'elaborazione di immagini» (Foundalis, 2006, p. 20), è anche inevitabile che un qualche meccanismo in grado di attuare processi di percezione visiva sia necessario all'operatività in questo dominio, meccanismo che, dunque, conferisce un'apertura in un certo senso non vincolata agli elementi che possono essere oggetto di rappresentazione da parte del programma. È per questo che, «il dominio dei Problemi di Bongard include alcuni elementi che

appaiono essere centrali nella cognizione umana; [...esso] è illusoriamente percepito come un microdominio, e non dovrebbe essere inteso come limitato da rigidi confini. Nel dominio dei Problemi di Bongard *la mente è il limite*» (*ivi*, p.21 [enfasi mia]).

Queste parole costituiscono il culmine della complessità nella scelta dei microdomini come universi di azione di programmi sviluppati all'interno dell'approccio subcognitivo, i quali ricercano un punto di equilibrio fra non limitatezza e duttilità da una parte e stabilità della rappresentazione dall'altra. Con PHAEACO è possibile constatare come tale obiettivo è tanto più raggiunto quanto più si riesce ad allargare la forbice fra processi cognitivi di alto livello e percettivi di basso livello, laddove i due aspetti non vanno visti in modo separato ma in continuità, con il fine ultimo di stabilire i limiti effettivi di questa capacità *mentale* considerata nel suo complesso.

Conseguenza di questa impostazione specifica è che l'architettura computazionale del sistema è, pur con alcune differente particolari, la stessa dei modelli precedenti e richiama in qualche modo la tripartizione alla base della teoria che ne costituisce la matrice. Il ciclo di interazione principale è tra uno Spazio di Lavoro e una Memoria a Lungo Termine, in cui sono immagazzinati i concetti permanenti. Il sistema comincia con processi *bottom up* di esplorazione dei riquadri del problema e prosegue facendo intervenire processi sempre più astratti. L'elaborazione delle immagini viene suddivisa in una sequenza di processi gerarchici ascendenti che lavorano sui *pixel* con l'obiettivo di costruire una rappresentazione delle figure esperite. Tale rappresentazione è costruita nello Spazio di Lavoro attraverso una serie di grafi ad albero, che hanno come nodo radice il nodo-riquadro e sotto-nodi quelli che rappresentano gli oggetti percepiti e, a un livello ancora più basso, le loro caratteristiche, ulteriormente scomponibili. Gli archi rappresentano relazioni di appartenenza dal basso vero l'alto. Così, se la figura percepita è un triangolo, il nodo corrispondente sarà inserito nella struttura gerarchica che comprende superiormente il nodo riquadro e inferiormente le sue componenti, ad esempio i lati, i quali avranno a loro volta nodi inferiori che ne indicano la lunghezza, l'orientamento, e così via. Ogni nodo è espresso da una serie di valori statistici tra i quali sono compresi il numero delle osservazioni, la media fra i valori delle osservazioni, la media della variazione, la somma dei quadrati. Ciò è conforme a una rappresentazione flessibile in grado di far convergere la presenza di differenti esempi verso un valore di stabilità che rappresenta la loro media. La struttura ad albero in realtà non ha la forma di un grafo aciclico perché è possibile che da diversi nodi parta un collegamento a un identico nodo sottoposto. Un esempio è il caso del nodo che esprime la numerosità di una caratteristica come il numero dei lati, il quale riporterà il valore corrispondente e su cui convergeranno tutti i nodi che rappresentano i lati.

L'indeterminatezza del grado di dettaglio degli elementi percepibili potrebbe far sì che il processo di costruzione, che non è limitato, procedesse senza fine, fatto implausibile dal punto di vista cognitivo. Per risolvere questo problema, ad ogni nodo del grafo è connessa una variabile che esprime l'attivazione corrispondente. Nel momento in cui la somma delle attivazioni dei nodi sottoposti trasmesse al nodo radice raggiunge un certo valore di soglia la costruzione della

rappresentazione termina. La struttura che ne deriva è gerarchica e i tipi di nodi che sono utilizzabili possono essere raggruppati secondo tre categorie: nodi oggetto, nodi caratteristica, nodi numerosità. In particolare, i nodi caratteristica che costituiscono la maggior parte della struttura ad albero sono di diverso tipo: punti, vertici, angoli, contatti, concavità, ma anche, come si è visto prima, nodi che esprimono caratteristiche costitutive interne (tessitura, riempimento) o relazioni (interiorità, lunghezza, uguaglianza). In particolare, questi ultimi, conformemente ai modelli precedenti, sono quelli su cui si basano i raggruppamenti e che inviano informazione "di alto livello" alla rete concettuale, facendo attivare i nodi della rete semantica permanente più astratti. In definitiva, tale processo costruttivo è basato su un numero limitato di primitive percettive visive, che Foundalis suppone essere di poche centinaia, sulla base delle buone prestazioni rappresentative di PHAEACO, e oltre le quali cominciano le ripetizioni[68].

L'aspetto più interessante di questa costruzione della rappresentazione risiede nel fatto che essa si espliciti, più ancora che nei modelli visti in precedenza, come gerarchia di tipi concettuali disposta su più livelli e che tale struttura ad albero rispecchi, dall'alto verso il basso, la *descrizione intensionale* di un oggetto, mentre i processi che inviano attivazione ai nodi della rete semantica, sia a partire da singoli nodi relazioni che puntano su più caratteristiche all'interno di uno stesso albero (è il caso, ad esempio, del nodo "uguaglianza"), sia nel caso di nodi caratteristiche uguali appartenenti ad alberi diversi, possono essere considerati la controparte della *descrizione estensionale* degli oggetti percepiti. Tale intreccio, che avviene sempre secondo le dinamiche probabilistiche della scansione parallela a schiera e dietro il superamento di valori di soglia, dispiega un sofisticato meccanismo non solo per la rappresentazione delle due dimensioni attraverso cui tradizionalmente viene definito un concetto, ma anche per il loro utilizzo dinamico in sede di percezione della situazione. Dal punto di vista teorico è stata avanzata, infatti, l'ipotesi (Linhares, 2000) che un sistema in grado di operare nel dominio dei Problemi di Bongard debba incorporare la possibilità di istituire la relazione percettiva secondo uno schema "molti-molti", ovvero, mentre ad ogni descrizione devono corrispondere molteplici segmentazioni dell'immagine elaborata, ogni segmentazione deve essere passibile di molteplici descrizioni. Il rapporto fra intensionale ed estensionale assume perciò un connotato variabile e flessibile, strettamente dipendente dal contesto dell'elaborazione, che procede a predisporre il punto di vista migliore a seconda delle esigenze attuali del sistema[69].

Tale gioco di rimandi trova il suo vincolo "superiore", cioè a livello cognitivo, nel compimento del processo di *pattern matching* attraverso la comparazione delle descrizioni, che sono le strutture

[68] Il convezionalismo nella metodologia di individuazione delle primitive visive è dichiarato esplicitamente dall'autore: «L'ipotesi fatta in questa tesi è che l'insieme delle primitive visive che possono essere espresse nei Problemi di Bongard [un mondo di figure bidimensionali] è grande – presumibilmente dell'ordine di qualche centinaia. [Tuttavia] la decisione se un dato tratto costituisca o no un primitivo è soggettiva» (Foundalis, 2006, p. 209). D'altra parte, ciò che sembra appartenere allo sviluppo del progetto non è la loro elencazione esaustiva, ma la dimostrazione che essi siano necessari all'espletamento di alcune attività cognitive basilari.

[69] Si rimanda a Linhares (2000) per un discussione filosofica sull'ontologia del dominio definito dai Problemi di Bongard.

gerarchiche ad albero sopra descritte. Tali strutture, infatti, possono essere considerate alla stregua di *esemplari*, la cui somiglianza è colta da uno specifico algoritmo del sistema basato sulla comparazione dei rispettivi livelli gerarchici. Il ritrovamento di caratteristiche simili fa sì che PHAEACO consideri un esemplare simile a un *pattern* (inclusione categoriale) e aumenti la stabilità di questo, incrementando il numero degli esemplari da cui lo ha ricavato e facendo la media fra le caratteristiche del nuovo esemplare e quelle che esprimono il *pattern* in quanto media degli esemplari già "inglobati". Un algoritmo di questo tipo è chiaramente anche in grado di eseguire un'operazione più basilare rispetto di quella del inclusione nel *pattern*, ma di un'importanza fondamentale, cioè la formazione di nuovi *pattern* a partire da due o più esemplari confrontati e scoperti come simili, al solito attraverso una funzione che calcola la media delle loro caratteristiche accoppiabili (si ricordi che le caratteristiche sono in realtà liste di valori numerici che esprimono parametri statistici).

Al di là degli aspetti più tecnici, l'implementazione di un meccanismo di questo tipo (costruzione di strutture ad albero + algoritmo di *pattern matching*) costituisce il punto di congiunzione fra processi di riconoscimento categoriale e di costruzione di analogie. Si può affermare, perciò, che uno degli obiettivi di PHAEACO consiste proprio nell'impostare in maniera effettiva la questione della capacità di fare analogie *come* attività di *pattern matching*[70] e questo a un livello di dettaglio che permette di racchiudere sotto un'unica prospettiva riconoscimento categoriale e processo di creazione di analoge attraverso l'impiego della nozione, complessa e sfaccettata dal punto di vista cognitivo, di *pattern*. A supporto teorico di questa concezione c'è la propensione a favore di una tesi che mescola *realismo ontologico* e *verticismo* (nel sistema che percepisce) delle funzioni percettive in una dimensione evoluzionistica, come è evidente nelle seguenti parole:

> La nostra abilità nel fare analogie (come apice), o *pattern matching* (come aspetto di base) – qualsiasi nome gli si voglia dare – consiste nell'abilità fondamentale delle creature cognitive di percepire il mondo e rivestirlo di senso, assegnando ciascun oggetto a una categoria conosciuta; di percepire le categorie attraverso l'esposizione a oggetti sufficientemente simili; e anche di percepire gli oggetti stessi, che è un prerequisito della categorizzazione. Come mettiamo in atto il vedere "oggetti" nel mondo, piuttosto che casuali collezioni di "pixel" inviati alla nostra corteccia visiva attraverso le aste e i coni della nostra retina? Lo facciamo perché alcune collezioni di "pixel", a causa della vicinanza spaziale (come in un insieme di punti), o alla vicinanza dovuta ad altre caratteristiche (colore, tessitura, ecc.) sembrano "stare insieme". Formando gruppi di ciò che sembra stare insieme, percepiamo gli oggetti.
>
> Si noti che l'uso di "noi" nel paragrafo precedente non implica che gli oggetti sono soltanto artefatti della cognizione. Gli oggetti devono esistere nel mondo; gli animali semplicemente evolvono nella loro percezione. Il presente lavoro può essere visto come una dimostrazione di esistenza delle proposizione

[70] Uno degli slogan presenti nel lavoro di Foundalis è: «Il *pattern matching* come nucleo centrale del fare analogie» (Foudalis, 2006, pp. 239 e sgg.), il quale costituisce un richiamo esplicito alla teorie esposte i Hofstadter (2001).

> che le menti non sono necessarie per percepire e così verificare l'esistenza di oggetti. Dopo tutto, anche PHAEACO può percepirli. (Foundalis, 2006, p. 242-243)

Il passo è molto denso e riassume molti degli aspetti visti nei precedenti modelli: la prospettiva simulativa relativa ai fenomeni mentali e non cerebrali, l'unificazione dei procedimenti di riconoscimento e di creazione di analogie e, soprattutto, le molteplici sfaccettature di cui deve essere dotata la conoscenza di un programma per poter produrre performance *significativamente* valide e che, concordemente a una certa visione filosofica, possono essere addotte a dimostrazioni di un realismo ontologico costruttivista dal punto di vista percettivo. Infatti, come già si era visto in TABLETOP in merito alla percezione della disposizione di oggetti tra loro collegati secondo una serie di relazioni categoriali, il programma deve poter disporre sia di concetti che esprimano relazioni fra categorie, sia di concetti che esprimano relazioni spaziali, sia di concetti così astratti che permettano di esprimere meta-relazioni fra quelle menzionate, anche in caso di disomogenità, così che sia possibile vedere, se necessario, l'uguaglianza fra due figure analoga all'uguaglianza fra due tessiture, cioè, come membri appartenenti entrambi a una stessa relazione astratta. Allo stesso tempo, il livello di dettaglio della strutturazione delle raffigurazione deve poter raggiungere in PHAEACO un livello di dettaglio molto elevato, come era stato in LETTER SPIRIT e in LETTER SPIRIT 2 per poter procedere a un adeguato riconoscimento e alla creazione di un stile quanto più omogeneo i dettagli delle lettere permettono.

Tale conoscenza è implementata in PHAEACO attraverso una memoria di concetti permanenti molto complessa, della quale vale la pena considerare brevemente alcuni aspetti. Di fatto, essa è costruita per replicare le strutture costruite nello Spazio di Lavoro. Perciò, più che di concetti si parla di «strutture nucleo concettuali» (*ivi*, p. 250), composte di un nodo centrale che rappresenta un oggetto e di nodi collegati ad esso che esprimono le sue caratteristiche. In tal modo viene facilitata l'attivazione di un concetto a partire dalle sue caratteristiche, ma anche l'operazione inversa di attivare le caratteristiche sulla spinta del nodo che rappresenta l'oggetto. Inoltre, i nodi caratteristica convergono sul nodo che rappresenta il nodo tipo della caratteristica (ad esempio, il nodo ideale – platonico – "*vertice*" cui sono connessi tutti i nodi vertice che fanno parte delle strutture nucleo. Questo meccanismo serve a istituire le associazioni fra queste ultime.

La rete è in grado di simulare anche funzioni di alto livello connesse con i processi mnemonici. Ad esempio, l'attivazione di un nodo relazione che etichetta (come in TABLETOP) una connessione fra due concetti causa il loro avvicinamento simulando il fenomeno dell'associazione. Il processo inverso di diminuzione dell'attivazione non sfocia nel ritorno alle condizioni iniziali, ma perviene al ristabilimento di una distanza minore di quella iniziale, simulando in tal modo l'andamento temporalmente determinato e selettivo dell'oblio. Ad ogni quantità di attivazione positiva corrisponde, infatti, un nuovo avvicinamento e un successivo minore distanziamento, a meno che ciò non avvenga su cicli di tempo molto lunghi. Esiste, poi, una collezione di nodi

"indessicali" che mettono in collegamento lo spazio di lavoro con la memoria permanente[71]. Infine, e questo è forse l'aspetto più rilevante, alla rete concettuale di PHAEACO possono essere aggiunti nuovi nodi che rappresentano nuove strutture nucleo concettuali. Si può dunque affermare che, a differenza dei modelli precedenti, e in conseguenza delle necessità di fronteggiare un dominio visivo virtualmente indeterminato attraverso un bagaglio di conoscenza *non totalmente pre-programmabile*, nel sistema è implementata una capacità di *learning* articolata su differenti piani. Il programma apprende nel senso che è in grado di *istituire associazioni* fra i concetti (modificandone in maniera graduale le distanze), *dimenticare* l'informazione irrilevante, *arricchire* concetti esistenti, *creare* nuove strutture concettuali[72].

Si è detto che l'elaborazione del programma procede in maniera *bottom up*. Questo è vero soltanto nelle fasi iniziali dell'elaborazione delle immagini, in maniera conforme a ciò che si è visto negli altri modelli. Ben presto, infatti, intervengono le microprocedure immesse dall'attivazione dei concetti della rete semantica, che servono a guidare l'elaborazione verso una comprensione in termini più astratti, cioè di alto livello e a una profondità concettuale maggiore. In effetti, molti dei problemi analogici vengono risolti attraverso la "giusta" corrispondenza di concetti complessi, che riguardano il livello cosciente, e, dal punto di vista dell'ontologia del dominio, dal nodo centrale che rappresenta gli oggetti, mentre le caratteristiche vengono lasciate in disparte una volta che l'elaborazione visiva è compiuta. Tuttavia, anche questo processo non è deterministico. È possibile, ad esempio, che sia necessario un esame "dall'alto" a un maggiore grado di dettaglio, che provoca un ritorno alla considerazione delle caratteristiche.

Tutto ciò si riflette nel modo in cui il programma cerca di arrivare a una soluzione. I *modi operandi* che adopera sono tre e vengono utilizzati gerarchicamente. Il primo è definito "circuitale" (*hardwired*) e vuole essere analogo ai processi per cui certi tipi di riconoscimento, in questo caso visuale, dipendono strettamente dai meccanismi neuronali implicati dal meccanismo sensoriale, cioè dai meccanismi di basso livello della percezione (visiva). Ne sono esempio tutti quei problemi che dipendono dalla presenza di figure con diversa tessitura o colorate e non colorate. Il secondo meccanismo è chiamato "olistico" ed è connesso al ritrovamento di una caratteristica comune fra le rappresentazioni costruite, la cui entità varia considerevolmente fra riquadri di destra e di sinistra. Si può pensare, ad esempio, a una differenza relativa alle aree, grandi quelle delle figure di destra, piccole per quelle di sinistra. Essa sarà notata abbastanza in fretta dal sistema nel processo di *pattern matching*. Se i primi due meccanismi non funzionano nel proporre ipotesi di soluzione, anche dopo un processo di produzione e controllo di più tentativi di soluzione "immediata",

[71] Un meccanismo molto simile per il recupero dell'informazione immagazzinata in memoria è stato suggerito, a livello teorico, anche da Minsky (1986)

[72] Eventualmente anche attraverso una parte del sistema chiamata "Mentore", in cui l'utente esterno può disegnare e assegnare un nome a nuove figure, o anche soltanto procedere al battesimo di una struttura percepita nella fase di elaborazione delle immagini. Più specificamente, l'interfaccia del programmamostra un'area riservata al disegno e una dedicata alla proposizione che descrive la figura (ad esempio, "cerchio nel pentagono"). Dopo diverse ripetizioni il programma impara ad associare, sulla base delle relazioni e dei concetti che già possiede (inclusione, cerchio), il nuovo nome alla nuova figura (pentagono) (cfr. Foudalis, 2006, pp. 102-106).

interviene il terzo tipo di elaborazione, definito "analitico", in base al quale il sistema prende in esame riquadro per riquadro (ovvero le strutture ad albero che rappresentano le figure dei riquadri) tentando di trovare caratteristiche simili in due o più di essi, fino alla enucleazione di un'ipotesi.

Senza entrare troppo nei dettagli, si deve sottolineare il fatto che questi tre tipi di elaborazione non corrispondono in realtà a tre moduli separati, ma potrebbero essere visti come tre differenti euristiche di soluzione, attuate dal consueto apparato delle microprocedure unitamente all'attività di *pattern matching* che costituisce, come si è visto, l'essenza del programma. In tutte e tre i casi, infatti, le operazioni effettive consistono nel confronto e nell'allineamento di caratteristiche, e sono compiute dalle microprocedure esplorative degli alberi che rappresentanto le strutture e dagli algoritmi che unificano le strutture formando *pattern* attraverso il calcolo della media dei valori delle caratteristiche stesse. Se queste riguardano aspetti percettivi di basso livello, si avrà il successo della strategia circuitale; se riguardano il rinvenimento di un qualche tratto omogeneo si avrà la riuscita della strategia olistica; se il processo invece porta ad accoppiamenti frammentati e alla creazione di *pattern* parziali, l'andamento del programma è da considerare analitico. I tempi di esecuzione si allungano concordemente al numero dei *pattern* formati, che nel terzo caso può essere superiore ad uno anche in riferimento al confronto di due medesime figure.

L'insistenza su questi aspetti acquista senso se si considera PHAEACO dal punto di vista della sua portata simulativa. Occorre chiedersi, infatti, che cosa effettivamente intende simulare questo programma. La risposta è principalmente una: il *pattern matching*. Infatti, non si può dire che la simulazione della strategia circuitale trovi effettivo riscontro nel modo in cui funzionano i processi cerebrali. Tuttavia, i processi di comparazione di alto livello, sul versante cognitivo per così dire, e dipendenti dai meccanismi di *pattern matching* possono essere considerati il candidato principale dell'intento simulativo del programma. Questo appare più chiaro se si considera che, come fa notare Foundalis, si è scelto di implementare il *pattern matching* attraverso algoritmi di generalizzazione a partire da esempi, piuttosto che con tecniche di *clustering*, basate sulla scansione di ampie masse di dati sottoposte al controllo di un'unità che applica dall'alto etichette classificatorie (*ivi*, pp. 227 e sgg.). La scelta di operare a partire da esempi, non dimenticando che sono "percepiti in maniera diretta" dal sistema, e l'insieme dei meccanismi stratificati di strutturazione e messa a confronto dell'input permettono al programma di affrontare problemi in cui è richiesta l'individuazione di relazioni di somiglianza (e non solo di identità) all'interno di un compito, quello della risoluzione dei Problemi di Bongard[73], che consiste in ultima analisi nell'individuare meta-relazioni di differenza.

In definitiva, questo sistema testimonia che la scelta di operare in un dominio visivo in cui le forme in gioco non sono predefinite implica la necessità di ricorrere a un modulo, interno al modello, che attui l'elaborazione delle immagini, anche se il livello effettivo della simulazione

[73] In questo caso si può parlare di pre-programmazione, perché il programma sa già in partenza che tipo di problema dovrà affrontare: confrontare i riquadri di destra, ecc...

rimane quello cognitivo. Foundalis propone di distinguere i processi implicati dall'affrontare i Problemi di Bongard in due livelli, uno definito "retinico", l'altro"cognitivo", sottolineandone la loro necessaria interattività dal punto di vista simulativo. Detto in altro modo, mentre lo scopo del programma non è quello di implementare una versione particolare di elaboratore di immagini, questo diventa rilevante nella misura in cui è un «processamento *cognitivamente interessante* dell'input, che comincia ad livello molto basso (grezzo)» (*ivi*, p. 71). All'opposto ci sono i processi cognitivi come la formazione di *pattern* visivi, il *pattern-matching* e l'immagazzinamento e ritrovamento all'interno della memoria a lungo termine dei *pattern* formati. La differenza fondamentale viene fatta consistere nel fatto che mentre il livello retinico «è piuttosto concorde all'elaborazione dell'input visivo nella retina e nella corteccia visiva, sebbene non aspiri a essere modello di quei *moduli* cerebrali al livello neurofisiologico, [il livello cognitivo] giunge *molto più vicino a modellare i processi psicologici umani*, ed è il livello al quale è impiegato uno schema di rappresentazione concettuale» (*ivi*, p. 72 [enfasi mia]). L'architettura del programma è, d'altra parte, strutturata in modo che seppure i due livelli siano in qualche maniera non permeabili l'un l'altro quanto ad attività, nel senso che nessuno sa quello che l'altro sta facendo, essi si influenzano a vicenda fin dalle prime fasi di elaborazione, il primo fornendo progressivamente i risultati, anche parziali, del processo di elaborazione visiva, il secondo fornendo di ritorno una direzione alle attività visive in base all'evoluzione dell'attivazione delle sue strutture concettuali, che subiscono un continuo processo di esplorazione e "confronto e allineamento" (*matching*).

Tale prospettiva modulare relativa ai processi mentali, che ha incontrato il favore di numerosi studiosi negli ultimi anni, sia per quanto riguarda un approccio esclusivamente mentale, sia da parte di neurofisiologi e neuropsicologi, trova un ampio riscontro nei modelli FARG che abbiamo considerato, per il fatto di muoversi su più livelli che comunicano l'uno all'altro attraverso i risultati delle elaborazioni intra-livello, ma rivestiti di una sorta di impermeabilità eleborativa inter-livello. Tutto ciò è particolarmente evidente in PHAEACO, dove la continua influenza fra livelli alti e livelli basi è solo *indirettamente* un'influenza che agisce sul processo generale di elaborazione, nel senso che non esiste un'unità di controllo centrale, al di là dell'algoritmo che regola la scansione parallela a schiera delle microprocedure e il loro andamento probabilistico[74].

Un'impostazione di questo tipo porta ad alcune conclusioni. Se, infatti, la simulazione riguarda effettivamente il livello cognitivo e non quelli più "bassi", l'allargamento dello spettro delle funzioni cognitivo-percettive richieste dal dominio, che abbiamo evidenziato all'inizio di questo paragrafo, conduce alla riproposizione del tema caro al funzionalismo e alle discipline simulative della *realizzabilità multipla*, tuttavia in un'accezione più articolata che potremmo definire "morbida" o "indebolita". Mentre la tesi della realizzabilità multipla sostiene che le attività superiori del pensiero possono essere svolte da supporti di verso tipo, lasciando non specificato il

[74] «L'"elaborazione cognitiva"di PHAEACO, non comprende alcun tipo di ricerca strutturata ad albero in uno spazio crescente dal punto di vista combinatorio» (Foundalis, 2006, p. 71).

modo in cui esse si connettano ai loro supporti, una visione morbida della tesi, come quella presentata da Foundalis (fig. 1) e tipica degli approcci emergentisti che rientrano nella tipologia subcognitiva, implica:

- la completa identificazione del livello superiore, i fenomeni mentali, in sistemi biologici e artificiali (simulativi), che costituisce anche l'assunzione forte della tesi della realizzabilità multipla;

- una separazione netta fra processi di basso livello su supporti differenti, la cui simulazione non è prevista come obiettivo di questo approccio simulativo, soprattutto per quanto riguarda il livello dei neuroni e quello dei bit;

- una parziale crescente sovrapposizione di livelli, che mette in corrispondenza dal punto di vista funzionale, le basi dei processi di pensiero superiori.

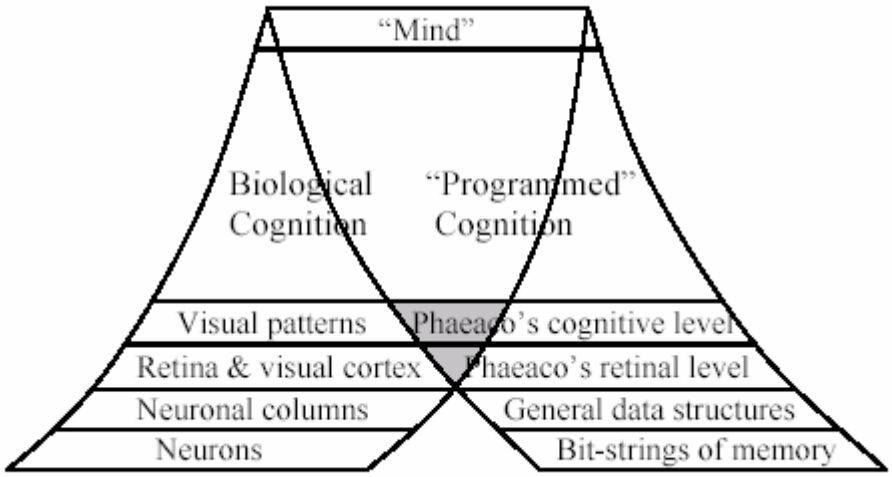

Fig. 3.6 - La crescente sovrapposizione dei livelli (tratto da Foundalis (2006, p. 72))

Lo schema illustrato nella figura 3.6 espone, perciò, nel dettaglio la tesi generale alla base di tutto l'approccio simulativo esaminato. Il livello più astratto, la "mente", sembra potersi rintracciare soltanto a livello delle prestazioni del programma, almeno in quelle in cui è in grado di fornire una risposta e soltanto nella misura in cui rispecchiano quelle di soggetti umani sottoposti ad esperimenti con i Problemi di Bongard. Con il diminuire dell'astrazione, la sovrapposizione diventa parziale. Ciò sembra spingere, ovviamente, verso un'interpretazione del tutto funzionale degli aspetti coincidenti. In altri termini, poiché la differenza fra "cognizione biologica" e "cognizione programmata", che sostituisce in Foundalis la dicitura "Intelligenza Artificiale", inizia già sotto il

grado più alto, ciò può essere considerato come indice del fatto che è inevitabile che a ogni livello ci sia una *quantità di simulazione limitata*, la quale, se da una parte è costitutiva della definizione della pratica stessa del "fare un modello di qualcosa", dall'altra equivale ad affermare che una simulazione totale non è neppure necessaria al fine di ottenere la simulazione dei *fenomeni mentali coscienti* che esprimono il grado massimo di astrazione: «la somiglianza aumenta in proporzione al grado di astrazione» (*ivi*, p. 73). Ciò che ancora una volta risulta rilevante è la funzione svolta dai meccanismi ai livelli intermedi.

Tale schema, dunque, più che indicare gli effettivi elementi di corrispondenza, può essere considerato una sorta di indicazione metodologica sul fatto che tali elementi di corrispondenza *devono esserci* e la spiegazione dei fenomeni ad essi collegati nella parte superiore *deve* implicare una discesa graduale di livello, se veramente si vuole considerare la nozione di *emergenza* valida dal punto di vista esplicativo. La conferma di questa indicazione sta nella realizzazione di sistemi in cui sia evidente e facilmente rintracciabile, cioè *esplicitata*, la costruzione basata su livelli gerarchici, nonché appaiano evidenti i modi in cui i livelli sono collegati e chiara e ben definita l'esatta natura delle funzioni simulata ad ogni livello. L'architettura di PHAEACO gode di questi requisiti, implementando un'ipotesi sugli elementi di corrispondenza dei livelli intermedi, la quale riguarda i meccanismi di *pattern matching* e oltre la quale l'unico aspetto di conferma non può arrivare che dal confronto delle prestazioni fra uomo e macchina.

Un'ultima considerazione in questo discorso va fatta in merito a un modello ancora in fase di sviluppo e volto alla simulazione del ragionamento in ambito geometrico. Per ciò che interessa i fini del nostro discorso, questo modello, denominato temporaneamente TRI-CYCLE[75], è pensato come uno scopritore di teoremi geometrici su figure come triangoli e cerchi. Il sistema presenta due parti: un'interfaccia e un modello cognitivo vero e proprio. Il dominio in cui opera è quello di una parte della geometria ed è, quindi, ancora una volta molto circoscritto. Un aspetto particolarmente interessante è che, seppure non si presenti come un modello simulativo di creazione di analogie, questo programma riprende il discorso lasciato in sospeso da PHAEACO per quanto riguarda l'elaborazione delle immagini. Infatti, l'obiettivo principale del progetto è lo sviluppo di un sistema in grado di avere una visione *human-like* della geometria. In particolare la conoscenza del programma, racchiusa nella sua memoria a lungo termine, è prevista contenere una serie di concetti geometrici che potrebbero essere posseduti anche da un essere umano e impiegati per l'analisi di una figura. In maniera conforme a questa impostazione, il programma non saprà l'esatta lunghezza di un lato, ma che esso è un segmento verticale, che appartiene a una figura con un "alto", un "basso", un "dentro", un "fuori", e così via.

Tale scelta operativa vuole essere un tentativo di simulare gli aspetti intuitivi legati alla scoperta in geometria. Perciò, da una parte come in PHAEACO, TRI-CYCLE dovrebbe essere in grado di

[75] Devo a Francisco Lara-Dammer, che sta lavorando a questo programma, le informazioni che ho raccolto in merito.

costruire rappresentazione di figure (triangoli, cerchi, punti, linee e segmenti) non solo sulla base delle loro caratteristiche "intrinseche" (numero dei lati o degli angoli), ma anche dal punto di vista delle loro reciproche relazioni. Dall'altra, diversamente che in PHAEACO, un'elaborazione delle immagini così condotta diventa un modo per ritrovare "elementi di sorpresa" che conducano all'individuazione di qualche teorema geometrico particolare. Basandosi sull'attivazione concettuale generata dall'elaborazione delle immagini, dunque, il programma dovrebbe essere in grado di ricavare teoremi non per via dimostrativa, ma ancora una volta sulla base dell'enucleazione di strutture concettuali *spazialmente* significative.

Questo lascia intravedere un'ambiguità, che forse verrà risolta negli sviluppi successivi del progetto. Essa riguarda il tipo di conoscenza che viene impiegata per ottenere la scoperta di un teorema e che può essere riassunta nella seguente domanda: quanta di essa deve riguardare la geometria e quanta un nucleo concettuale più astratto? Saranno presenti nella base semantica sia concetti riguardanti relazioni geometriche, come, ad esempio, la simmetria; sia concetti riguardanti relazioni più astratte come quella di somiglianza, che già in PHAEACO abbiamo visto sostituire la relazione di identità in quanto giudicata psicologicamente più plausibile nei processi di scoperta ed ottenuta attraverso gli algoritmi di *pattern formation* e *pattern matching*; sia, infine una serie di conoscenze relative a proprietà geometriche più specifiche che permettano al programma di riconoscere come importanti certe scoperte nello spazio percettivo. Tuttavia, con una differenza. Mentre in PHAEACO tali procedimenti coinvolgevano livelli inferiori rispetto a quello più astratto dell'elaborazione cosciente (come è naturale che sia, se il tentativo è quello di cogliere processi di categorizzazione attraverso esempi: se il cervello compie calcoli nel farlo, noi non ne siamo consapevoli), in TRI-CYCLE, tali relazioni devono entrare a far parte del bagaglio di conoscenza esplicita sulla base del quale è possibile non tanto il ritrovamento della soluzione, quanto piuttosto il complesso processo della scoperta, che implica la totale imprevedibilità del percorso scelto. Spesso, infatti, nel tentativo di trovare teoremi validi nel dominio della geometria siamo spinti alla costruzione di figure aggiuntive a quelle già presenti nella situazione, senza sapere effettivamente a cosa porteranno, ma operando attraverso un processo per tentativi ed errori, che coinvolge solo in parte processi di costruzione di analogie con conoscenze e situazioni precedenti, per lasciare il posto alla libera scoperta (permessa, si potrebbe dire hofstadterianamente, anche da una sensibilità estetica nei confronti della *forma*).

L'ambiguità di cui si diceva, viene rafforzata, più che risolta, dall'introduzione nel modello di un Piano Mentale (*Mental Plane*), in cui la strutturazione della situazione nello Spazio di Lavoro avviene attraverso uno schema dotato di coordinate spaziali. Al di là dei dettagli tecnici, il programma di Lara-Dammer può essere considerato, ancora più che PHAEACO, un modello volto a provare le tesi di Kosslyn (1980) sul ruolo svolto nella mente dalle immagini mentali, ovvero, da una rappresentazione concettuale della situazione secondo coordinate in uno spazio orientato. Il punto è, per ritornare alla domanda posta in precedenza, se tale piano o schermo mentale debba

contenere forme o concetti[76]. Lasciamo aperta tale questione. Vi ritorneremo da un punto di vista più generale nel prossimo capitolo.

In conclusione, possiamo affermare che il contributo del modello TRI-CYCLE, una volta terminato, sarà forse quello di mostrare che, se il tentativo di simulare la creatività comporta una buona dose di meccanismi per fare analogie, quello di simulare la scoperta (in questo caso in geometria), implica, in una certa misura e allo stesso tempo, anche la possibilità di indebolire i vincoli su cui ogni costruzione analogica si compie. La geometria come dominio fortemente strutturato e allo stesso tempo non prevedibile ripropone una tesi sottesa a tutti i modelli finora affrontati, e cioè che, se dal punto di vista del risultato, l'analogia sembra uno dei fenomeni centrali della cognizione, poiché è un buon candidato a essere un'esaustiva spiegazione della derivazione non formale della maggior parte delle conoscenze umane, dal punto di vista processuale, i meccanismi che mettono in atto il processo di creazione di analogie devono, di contro, comportare un'attenuazione proprio dei vincoli analogici, cioè su cui si basa la mappatura concettuale, e dunque l'immissione di casualità in misura tanto maggiore quanto più vasto, ricco e imprevedibile nei suoi elementi costitutivi è il dominio nel quale il processo di "scoperta del nuovo" viene messo in atto, così come è quello della geometria.

[76] Francisco Lara-Dammer (comunicazione personale).

Capitolo 4

SUBCOGNIZIONE, ANALOGIA E SIMBOLI ATTIVI: VERSO UNA NUOVA TEORIA DEI CONCETTI

4.1 Uno sguardo retrospettivo

Il capitolo conclusivo di questo lavoro è dedicato alla ripresa e alla valutazione dell'approccio allo studio dei processi di pensiero esposto nei precedenti capitoli, definito "subcognitivo", con un'attenzione particolare ai suoi presupposti e alle implicazioni sia dal punto di vista filosofico sia dal punto di vista dell'IA e delle scienze cognitive. Prendiamo lo spunto ripercorrendo la strada compiuta sin qui.

Nel primo capitolo sono state illustrate alcune questioni teoriche di cornice. Le domande che ci siamo posti riguardano uno dei problemi che hanno interessato più da vicino l'impresa dell'IA come disciplina volta allo studio dei molteplici fenomeni mentali in cui sono implicati il significato e la conoscenza, prendendo come punto di riferimento la nozione di intelligenza. È possibile che una macchina esibisca un comportamento intelligente? In che modo? Quali sono le condizioni che ci permettono di attribuire intelligenza a una macchina? Turing ha provato a fornire una risposta filosofica e pratica al tempo stesso a queste domande, proponendo il famoso gioco dell'imitazione e dando l'avvio, con ciò, alla sua controparte più tecnica, la teoria della simulazione in ambito cognitivo. Le idee di Turing sono state più volte riprese nel corso della seconda metà del Novecento, fino a essere criticate all'interno di un apparato teorico-filosofico, quello di Searle, volto alla sconfessione della ricerca in IA, in particolare quella simbolica, prevalente ai tempi in cui tale critica è stata formulata.

Il problema posto da Searle, indipendentemente dalle conclusioni cui arriva, è di ampio respiro. Comunemente viene inteso in questo modo: come è possibile che una macchina sia dotata di una capacità semantica? Una formulazione più appropriata potrebbe essere la seguente: se ascriviamo a un uomo una determinata capacità che chiamiamo "semantica", a quali condizioni possiamo considerare una macchina dotata della stessa capacità? La proposta che abbiamo fatto in quella sede è stata quella di volgere la nostra attenzione non tanto al modo in cui il linguaggio, inteso come sistema di segni, è dotato di significato, per poi andare a cercare ciò che manca ai linguaggi, cioè ai simboli, utilizzati da una macchina per potersi dire dotati di significato. Piuttosto, ci è sembrato più

giusto porre il problema dal punto di vista delle abilità percettive e rappresentative, e del modo in cui esse possono essere simulate su un sistema. La portata della domanda si è perciò allargata, così come si sono diversificate le risposte che le scienze cognitive hanno cercato di fornire in merito a tali questioni negli ultimi decenni. In che modo, sembra più giusto chiedersi infatti, è possibile costruire un sistema artificiale che simuli le capacità percettive umane "di alto livello", ovvero connaturate con i fenomeni cognitivi di categorizzazione, concettualizzazione, creazione di analogie, i quali sembrano implicare tanto l'apporto di informazione esterna al sistema, quanto la modificazione dell'informazione che detiene il sistema e la creazione di nuova informazione? Ciò implica necessariamente la comprensione del modo in cui il sistema viene approntato e non soltanto la stima della corrispondenza fra le sue risposte e quelle di un essere umano. Affrontare tali questioni ci è sembrata una strada obbligata per aggirare l'impossibilità di capire dove sta, per dirla con Leibniz, la vera natura della macchina percettiva, che si confronta con l'ambiente in cui agisce e che produce azioni dotate di significato in merito a specifici compiti, visto che, potendo osservare i pezzi della macchina che produce percezione, non siamo in grado di indicare con precisione il pezzo (fisico) che la produce.

Le risposte a queste domande chiamano in causa, necessariamente se si vuole una spiegazione, la comprensione dell'organizzazione funzionale dei meccanismi del pensiero, nonché una serie di quesiti sulla loro dipendenza/indipendenza dai meccanismi fisici che li rendono materialmente possibili, ma anche il riconoscimento del livello che si è disposti ad accettare come esplicativo. A tale proposito non si può prescindere dalla nozione di "funzionalismo", che, in modo onnicomprensivo (Cordeschi, 2002), è alla base delle discipline simulative, perché costituisce la condizione necessaria della loro valenza esplicativa.

L'approccio funzionale allo studio dei processi di pensiero che abbiamo introdotto nel secondo capitolo non trascura la possibilità che la spiegazione dei fenomeni mentali possa essere vista nella strutturazione di sistemi complessi, che riproducono aspetti di livello intermedio fra mente e cervello, perché avvengono al di sotto del livello dell'attenzione cosciente. La prospettiva individuata è stata così definita, dagli studiosi che l'hanno adottata, "subcognitiva" e riguarda il modo in cui la mente attua i processi percettivi compresi in uno spettro molto vasto che va dalla categorizzazione alla mappatura concettuale e alla produzione di contenuti di pensiero che sublimano, in complesse strutture concettuali, la distinzione fra categorie e processi.

Nel corso del terzo capitolo abbiamo passato in rassegna diversi modelli computazionali volti a questo scopo, sottolineando come *il problema della rappresentazione della conoscenza* in un modello vada di pari passo con *il problema del modo in cui quel modello è in grado di rappresentarsi la realtà*, cioè il dominio, in cui interagisce. Abbiamo anche visto come, nella prospettiva subcognitiva, la fusione di aspetti concettuali e materiale percepito è ciò che deve essere spiegato, quanto alle condizioni della sua attuazione, per capire *in che modo la mente arriva a*

dotarsi di contenuti significativi e allo stesso tempo a dotare i contenuti stessi di significato. I due processi non possono essere scissi, pena il ricadere nelle anomalie teoriche evidenziate da Searle.

In questo ultimo capitolo, riprenderemo il discorso da un punto di vista più generale, innanzitutto cercando di individuare le idee implicate dall'approccio che abbiamo preso in considerazione. Ricostruiremo poi la teoria del sistema mente/cervello che soggiace all'approccio subcognitivo. In seguito, considereremo l'aspetto più significativo della teoria computazionale espressa dall'approccio subcognitivo all'IA, il pensiero come risultato emergente dell'interazione dell'attività di micro-agenti. Ne vedremo i collegamenti con una teoria non subsimbolica dei processi di pensiero e concluderemo il discorso analizzando in che modo queste idee e la loro implementazione gettano luce sul problema, che riguarda anche la filosofia e la psicologia, della natura dei concetti, attraverso la proposta di una teoria che li vede come *analogie*.

Il discorso sarà introdotto affrontando in via preliminare alcune questioni epistemologiche riguardanti la ricerca nelle scienze cognitive e la difficoltà di valutare i suoi prodotti, questione che riguarda da vicino l'approccio preso in considerazione in questo lavoro proprio per il massiccio ricorso, a fini esplicativi, a un apparato funzionale apparentemente senza un esplicito riferimento simulativo, così come lo sono stati i neuroni per i nodi delle reti neurali e i termini e i costrutti linguistici per i programmi basati su formalismi logici di rappresentazione ed elaborazione della conoscenza.

4.2 Scienze, scienze della mente e scienze cognitive

Nel corso di questo lavoro sono stati toccati numerosi temi che riguardano problemi condivisi da più discipline interessate a spiegare i processi del pensiero e i fenomeni mentali. Tuttavia, parlare di mente in un'epoca in cui lo studio dei fenomeni che la riguardano è sempre più intrecciato con l'acquisizione di dati in merito al funzionamento del cervello appare quasi un'impresa anacronistica, di taglio storico più che teorico. I termini "mente" e "mentale" negli ultimi decenni hanno acquistato un sapore pre-scientifico e quasi *naïve* dal punto di vista filosofico. Eliminare dal discorso scientifico questa terminologia è un'impresa che, però, si è rivelata ardua anche a dispetto dell'enorme balzo in avanti compiuto dalle neuroscienze negli ultimi anni attraverso strumenti di indagine che permettono l'acquisizione di immagini in diretta del funzionamento del cervello, come nel caso della risonanza magnetica funzionale.

Tuttavia, se una svolta non c'è ancora stata è perché la "mente" resiste, non retoricamente, sia come termine descrittivo di un insieme di fenomeni, sia come campo di indagine privilegiato di alcune discipline che non potrebbero confrontarsi con la realtà che stanno analizzando se essa non comprendesse l'oggetto "mente", ben distinto dall'oggetto "cervello". Proviamo a immaginare cosa sarebbe la riflessione filosofica sul linguaggio e sul pensiero senza la possibilità di ricorrere al

mentale, ma anche come potrebbero prendere corpo numerose ricerche in differenti branche della psicologia senza il ricorso a un apparato teorico e terminologico che comprenda la possibilità di riferirsi a fenomeni specificamente mentali. O si pensi anche alle ricerche in una disciplina come l'IA, intesa in senso psicologistico, entrata ormai a far parte delle scienze cognitive, senza perdere tuttavia i suoi tratti peculiari di indagine simulativa dei processi di pensiero. Si potrebbe obiettare, però, che è solo questione di tempo e che nuove e più approfondite scoperte sul cervello mostreranno la superfluità dell'affidarsi a teorie che ancora comprendono un qualche riferimento alla mente, così come, ad esempio, la nascita di numerosi filoni biologistici e *neural-like* all'interno delle scienze cognitive sembra già indicare. Si potrebbe, cioè, sostenere che il riduzionismo fra mente e cervello, *da* mente *a* cervello, è lo stadio ultimo e inevitabile di ogni ricerca volta alla spiegazione definitiva dei fenomeni individuati come mentali.

Tuttavia, tutto porta a credere che la portata esplicativa di questi filoni di ricerca sarebbe estremamente impoverita senza un opportuno collegamento con un vocabolario che faccia uso di termini mentalistici e che la soluzione di questo particolare problema non sembra neppure all'orizzonte. Paradossalmente, è la stessa ricerca scientifica, per anni votata ad un abnegante riduzionismo, a mostrare i limiti di questa impostazione. Non è forse vero che i fondamenti ultimi della materia, le particelle subatomiche (del modello standard e di quello non standard) che appaiono "vivere" in un mondo complesso ma retto da leggi completamenti differenti da quelle del mondo in cui viviamo noi esseri umani, sono individuati in termini funzionali, essendo impossibile per definizione la loro identificazione concreta, oggettiva, materiale attraverso uniformi coordinate spazio-temporali? E cosa pensare delle recenti affermazioni del premio Nobel per la chimica Roald Hoffmann circa la natura *non* riduzionistica della propria disciplina alla fisica, fatto che al contrario viene dato per scontato dalla gran parte degli scienziati[1]?

In conformità a queste idee, la mente continua, dunque, a essere studiata come mente e il cervello come cervello. L'apparente dualismo ontologico cessa di dare fastidio nel momento in cui si riconoscono mente e cervello innanazitutto come due quadri concettuali. La coniugazione di questi due quadri sembra un'impresa molto meno difficile (appunto perché non impossibile) di quella del rapporto fra due sostanze disomogenee, espressioni come in Descartes non già di una fisica e di una anti-fisica, ma di due fisiche divergenti. Un'integrazione fra questi due universi separatamente indagabili sembra porsi sia come traguardo necessario sia, allo stesso tempo, come postulato della ricerca, ma non come indizio a favore o a riprova dell'inevitabilità del riduzionismo esplicativo (lasciando ancora da parte quello ontologico, che è un'altra questione ancora, relativa alle ontologie in gioco). Il problema sta nel modo in cui renderla effettiva e, dunque, in cui poter parlare in maniera sensata e adeguata di un apparentemente più opportuno e meno unilaterale "sistema mente-cervello" come *continuum* di livelli di fenomeni e di spiegazione di tali fenomeni.

[1] Si veda l'intervento di Rohald Hoffmann dal titolo "La bellezza della chimica" su *Il Sole-24 Ore. Domenica* del 7 gennaio 2007 (p. 33).

Il primo ostacolo che incontra la ricerca volta allo studio dei fenomeni mentali consiste senza dubbio nella difficoltà a presentarsi come una ricerca di stampo scientifico nel senso tradizionale del termine. Nel primo capitolo, abbiamo introdotto la questione in maniera provocatoria, ipotizzando che le discipline simulative sono un tipo di sperimentazione situato a metà fra gli esperimenti scientifici tradizionali e i così detti *Gedankenexperiment*. La provocazione sta nella disomogeneità dei due approcci, sia per quanto riguarda la metodologia impiegata, sia per quanto riguarda gli obiettivi e il modo di condurre le scelte teoriche che rendono l'esperimento dotato di significato effettivo. Mentre i primi hanno il compito di scoprire fatti o di confermare con i fatti le teorie entro cui vengono impostati ed eseguiti, i secondi mettono alla prova le teorie dal punto di vista della loro tenuta concettuale, forzando o a rivedere i concetti impiegati o a cambiare, cioè a valutare diversamente, la pratica sperimentale stessa in quanto costruzione dei processi di scoperta. Così nel campo dell'IA e delle scienze cognitive in alcuni casi ciò che è in gioco è la conferma o disconferma di una teoria, oppure la scelta di una fra più teorie rivali; in altri, invece, sono le metodologie impiegate ad essere oggetto di disputa; in altri ancora, la simulazione ha il compito di stabilire non quale fatto conferma un fenomeno, ma quale è effettivamente il fenomeno che viene indagato.

Tutto ciò ha sicuramente a che fare con il fatto che l'oggetto di studio delle scienze cognitive è visto in maniera diversa da ogni particolare "scienza cognitiva" e che le persone che si dedicano a questo campo di indagine, gli scienziati cognitivi, provengono da formazioni scientifiche e teoriche molto differenti, ognuna delle quali porta con sé un retroscena implicito di principi sulla natura della ricerca e dell'impresa scientifica molto diversi fra loro. Ad esempio, se un neurofisiologo individua una parte specifica della corteccia cerebrale come sede privilegiata dei processi di, poniamo, pianificazione e, allo stesso tempo, uno psicologo ricostruisce da una serie di esperimenti il modo in cui tali processi di pianificazione vengono attuati dagli esseri umani, come dovrà procedere la ricerca simulativa? Riprendendo la struttura della porzione di corteccia preposta al compito e simularla (vista la specificità delle aree cerebrali quanto a conformazione neuronale) o piuttosto focalizzando la sua attenzione sul modo in cui avvengono i processi di produzione di azioni pianificate cercando di simularli in un meccanismo astratto e generale di costruzione di tali processi? E inoltre, chi può candidarsi ad essere miglior giudice della riuscita dell'esperimento se non qualcuno disposto a vedere l'interrelazione fra questi due apporti, da una parte considerando il modo in cui gli esperimenti vengono condotti e, dall'altra, procedendo a un'integrazione concettuale fra dati, metodi e risultati?

Per tali ragioni, è stato proposto di non considerare, dal punto di vista epistemologico, l'IA come una scienza (Matteuzzi, 1995) poiché sprovvista di almeno due dei requisiti necessari all'unitarietà di ogni approccio scientifico: un universo univoco di riferimento e un linguaggio unitario di espressione. Tale affermazione sembra anche più giustificata, ancorché paradossale, se viene estesa alle scienze cognitive in generale, che salverebbero la loro scientificità definendosi in maniera

plurale, pur mantenendo un indefinito quanto generalmente riconosciuto obiettivo di fondo, quello della spiegazione del pensiero.

Affermazioni del genere non sono mancate neppure da parte di chi ha ideato e sviluppato l'approccio subcognitivo allo studio della mente. Si consideri, ad esempio, il seguente passo di Hofstadter, ripreso da un saggio sulla valutazione della ricerca in questo campo e che ci porta direttamente ad affrontare la "spinosa questione" di come considerare i risultati da essa conseguiti:

> [...] nell'ambito delle scienze cognitive/IA uno dei problemi più profondi è quello di riuscire a scoprire criteri universali che permettano di giudicare settori di ricerca. Il campo è molto confuso, giacché non sono poche le differenti pretese di validità, importanza e novità che vi si confrontano e competono, spesso parlando lingue del tutto diverse tra di loro. IA e scienze cognitive tentano di comprendere un fenomeno complesso al punto che ancora non si sa come giudicare le idee al riguardo.
>
> [...] In breve, l'insieme IA/scienze cognitive è un pazzo bazar, o almeno uno stravagante folle insieme di discipline. Lo spettro delle competenze scientifiche di chi vi opera è enorme, e i progetti sono i più disparati. (Hofstadter, 1995c, pp. 393).

4.3 Microprocedure e convalida cognitiva

Il problema di come valutare la ricerca in scienze cognitive interessa, dunque, anche l'approccio *subcognitivo*. Le cose sono, anzi, complicate dal fatto che ci si riferisce a un livello del sistema mente-cervello che non corrisponde a nessuno dei due estremi, quello fisico cerebrale e quello mentale cosciente. Come è possibile, dunque, valutare l'aspetto principale di questo approccio, cioè quello microprocedurale? Che valenza hanno le microprocedure dal punto di vista del computazionalismo inteso come teoria generale di spiegazione dei fenomeni mentali? Esse sono soltanto un dispositivo implementativo o è possibile pensare a una contropartita computazionale, in un senso che riguarda la visione della mente come computazionale, delle microprocedure? O è meglio identificarle attraverso un riferimento alla fisicità del cervello? O, ancora, l'ideale sarebbe trovare entrambi?

Per rispondere a queste domande occorre sgombrare il campo da alcuni equivoci che potrebbero essere in agguato. In primo luogo, qualsiasi strategia simulativa si adoperi, occorre non disconoscere la sua natura funzionale. Infatti, sia che si adoperino i metodi dell'IA simbolica tradizionale, sia che ci si muova all'interno di una visione connessionista, tutto è simulato, nulla è veramente ciò che, dal punto di vista della materialità, sostiene di essere. In altri termini, mentre appare intuitivo il funzionalismo di moduli computazionali, formati da conoscenza simbolica esplicita più regole di applicazione su tale conoscenza, va rivendicata la natura funzionale anche delle reti neurali come simulazione (semplificata del cervello), e questo indipendentemente dal fatto

che ogni implementazione di una qualsiasi simulazione può avvenire su dispositivi seriali, locali, discreti, quali sono i calcolatori tradizionali. Sembrerebbe, perciò, corretto pensare le microprocedure come simulazioni di apparati funzionali intermedi in una concezione funzionale di tutto il sistema mente-cervello, una sorta di ponte fra attività di basso livello (percettive e categorizzanti) e di alto livello (cognitive in senso classico), ponendo l'accento sul fatto che tutto il sistema va inteso come una *gerarchia di livelli di natura funzionale*, integrabili attraverso meccanismi di mediazione che sono anch'essi funzionali. Dunque, se è possibile affermare che il connessionismo si rifà alla neurofisiologia del cervello, simulandone i suoi costituenti (i neuroni) e l'IA di stampo classico tradizionale può in qualche forma essere ricondotta ai macro-apparati funzionali del cervello, secondo una visione che coincide in qualche modo con quella della neuropsicologia, a quale branca dello studio del cervello ci si può rivolgere per trovare il corrispondente cerebrale delle microprocedure?

French fa notare questo aspetto problematico sottolineando la difficoltà principale insita in una visione gerarchica funzionale del sistema mente-cervello: «la sola cosa da fare è convincere lo scettico che il tuo livello di indagine è quello appropriato. Se su questo ci può essere accordo, allora il grado di validità dei meccanismi di un particolare modello dipenderà da quanto bene il modello opera al livello concordato di indagine» (French, 1995, p. 147). L'accento è, dunque, sulla *performance* del modello, e questa prospettiva appare inevitabile, così come non è evitabile che la sua accettabilità si basi su un accordo teorico circa l'adeguatezza del livello simulato. D'altra parte, abbiamo già visto le considerazioni, aporetiche, di Rehling in merito all'esatta definizione di una controparte in termini cerebro-mentali per le microprocedure. Tuttavia, il discorso sulla convalida di questi modelli, se visto in riferimento alla loro *performance*, è stato a lungo affrontato da più punti di vista.

Hofstadter, ad esempio, rigetta l'idea che si possa ottenere una qualche utile indicazione in merito alle capacità indagate e simulate, nello specifico quella di fare analogie, attraverso una metodologia basata sulla media delle risposte a quesiti di analogia di un certo numero di soggetti umani:

> C'è qualcosa di profondamente sbagliato in tale idea: fare la media di un gruppo di menti di prim'ordine è un pasticcio come mescolare assieme gli ingredienti di diverse ricette famose nella speranza di ottenere, così, un piatto eccellente: puramente ridicolo! Le ricette famose e le menti brillanti sono uniche: farne la media le distrugge. [...] ogni stile cognitivo individuale svanisce. Allora, un obiettivo più ragionevole, per un modello del fare analogie, potrebbe essere quello di agire come una *particolare* mente creativa, o forse anche di riuscire a comportarsi come svariate menti creative, al variare di certi «parametri cognitivi» critici. (Hofstadter, 1995c, p. 386).

La presa di posizione contro la metodologia impiegata dalla psicologia sperimentale è netta. Una critica di tal genere è, peraltro, in linea con l'atteggiamento simulativo all'interno delle scienze

cognitive, il quale prende le mosse dalla nozione di simulazione[2], ma non affronta mai fino in fondo la sua problematicità. L'ipotesi di meccanismi microprocedurali è, in questa prospettiva, intermedia nel senso che costituisce un'alternativa alle metodologie sia psicologiche che simulative impiegate.

La metodologia della psicologia sperimentale, infatti, viene considerata essere troppo coinvolta e *ad hoc*, o al contrario falsamente generalizzante, nel descrivere i fenomeni cognitivi attraverso statistiche che rispecchiano la media delle prestazioni umane su semplici compiti, i quali hanno come scopo l'identificazione e lo studio di diversi effetti, come, per citarne alcuni, l'effetto *priming* o altri effetti associativi, o anche, per quanto riguarda attività cognitive "superiori", i fenomeni di deviazione dalla razionalità perfetta nei ragionamenti sulla utilità attesa studiati da Tversky e Kahneman. L'individuazione di tali fenomeni, inoltre, non corrisponde alla spiegazione dei loro meccanismi. Di contro, la scelta di operare attraverso resoconti introspettivi da ritrasporre nella formulazione di modelli computazionali teorici, tipica di una certa tradizione dell'IA (Newell, Simon, 1972) sembra troppo compromessa con il vecchio e difficile da giustificare introspezionismo, accusato di scarsa oggettività già alla fine del diciannovesimo secolo. La simulazione di aspetti macro-funzionali della mente, quali moduli dedicati a questa o a quella funzione cognitiva superiore e spesso basati sull'utilizzo di tecniche simbolico-formali, nel senso di sintatticamente manipolabili, viene tacciata di un eccessivo teoricismo, esplicativo ma difficilmente confermabile dal punto di vista dell'essere umano, vuoi per il problema del riferimento che concerne ogni apparato simbolico, vuoi per la non ostensibilità della mente stessa, se non per quanto riguarda i suoi prodotti (spesso simbolici, cioè linguistici o linguisticamente esprimibili). D'altra parte, infine, il ricorso massiccio alla metodologia connessionista, che risolve il problema dell'ostensibilità dell'oggetto simulato, la macchina-cervello, appare per assurdo tanto meno esplicativo, quanto più poggia sulla giustificazione di essere la "simulazione del giusto livello", cioè quello neuronale, perdendo un po' di questa limitazione nel momento in cui mette da parte proprio questo assunto epistemologico di fondo.

Sulla base di queste premesse, la proposta di Hofstadter di individuare i "parametri cognitivi critici" sembra una scappatoia, che va, però, valutata secondo la giusta angolazione. La proposta riguarda, come si è detto, il problema di valutare modelli che compiono analogie. La lista, incompleta, dei parametri suggeriti da Hofstadter (1995c, pp. 386-387) comprende una serie di capacità che vanno dall'attenzione prestata a elementi percepiti, alla velocità dei processi percettivi, al rinvenimento di relazioni di somiglianza, alle attivazioni concettuali connesse con queste operazioni, agli slittamenti operati, alla creazione di punti vista complessivi, cioè percezioni di alto livello, e alla competizione fra punti di vista alternativi e controfattuali. Questi parametri rendono attuabile l'analisi del *comportamento* dei modelli sviluppati, grazie alla possibilità di guardare il modello nel corso della sua elaborazione. Tutto questo è dovuto alle potenzialità delle

[2] Si vedano l'*incipit* del testo più che conosciuto sulla convocazione del seminario di Dartmouth agli albori dell'IA (McCarthy, Minsky, Rochester, Shannon, 1955), come anche le riflessioni pionieristiche di Craik sulla nozione di modello (Craik, 1943), che il *Proposal* di Dartmouth riprende esplicitamente.

microprocedure e, in particolar modo, della loro interazione complessa. Si può concludere che è l'elevato numero dei processi interagenti, più che la loro natura individuale effettiva, a determinare la valenza simulativa-esplicativa dei modelli subcognitivi, grazie alla possibilità di osservare il comportamento *interno* dei modelli, basato su un procedimento al tempo stesso meccanico-funzionale e dinamico-evolutivo. Infatti, i modelli subcognitivi:

> mirano a simulare la maniera in cui interagisce un numero molto grande di minuscoli meccanismi indipendenti che operano di concerto producendo un comportamento emergente di alto livello. Quando questi meccanismi differenti sono, per esempio, un centinaio, ognuno con una variabilità sua propria, anche se molto limitata, allora la loro interazione globale possiede un numero enorme di *gradi di libertà*. Quando ne siano coinvolti moltissimi, simultaneamente, l'insistere a considerare che i criteri di convalida di un comportamento così complesso possano essere identici a quelli usati per modelli di un meccanismo singolo crea solo confusione, dovuta all'abitudine e alla pratica. (*ivi*, p. 388 [enfasi mia])

Il problema del giusto livello di simulazione, nota Hofstadter, è largamente connesso con l'indeterminatezza della nozione di "strutture cerebrali" che, egli sottolinea, si è sostituita a quella di "meccanismi mentali" con la sempre maggiore attenzione riservata alle metodologie e ai principi del connessionismo. Infatti, quale può essere il giusto livello cui va esplorato il cervello, ovvero a quali strutture cerebrali bisogna fare riferimento se se ne possono individuare molte e diverse, quali, solo per fare alcuni esempi, il livello atomico, quello molecolare, quello delle cellule o delle parti cellulari (assoni, dendriti, sinapsi), quello di insiemi cellulari, quello di parti della corteccia, o addirittura un emisfero intero? Un discorso analogo può essere fatto per i "meccanismi mentali". Anche qui il giusto livello è da scegliere fra quello dei concetti semplici, delle catene associative, dei moduli di memoria o degli schemi contestuali come i *frame* e gli *script* (Hofstadter, 1996), ricordando che, se esiste la tentazione di considerare questi ultimi meccanismi come astrazioni teoriche o come dispositivi funzionali, non sembra implausibile, e non c'è alcuna impossibilità di principio, nel pensare che ad essi possa essere associata una controparte fisica in modo non esclusivamente riduzionista, come avviene generalmente quando si associa una particolare porzione del cervello ad una particolare funzione a seguito di esperimenti con elettrodi, metodo più vecchio, o di mappature ottenute con risonanze magnetiche funzionali, metodo di ultima generazione.

Naturalmente, come fa notare anche Hofstadter, strutture cerebrali e meccanismi mentali hanno una diversa caratterizzazione. Nel primo caso si tratta di strutture individuabili *fisicamente*, nel secondo di meccanismi qualificati dalla funzione che svolgono, la cui controparte fisica è, se non problematica, una sorta di traguardo della ricerca. Il fatto che esistano molti tipi di strutture cerebrali a vari livelli, indica l'ambiguità della nozione di "struttura cerebrale" se considerata dal punto di vista fisico. Posto che l'indagine di ognuno di questi livelli può portare a risultati fruttuosi ed essere oggetto di differenti campi scientifici di indagine, non è ben chiaro quale sia il livello del

“cervello”. Esso, infatti, diventa una nozione troppo vaga, una «categoria platonica, [che] rivela la nostra ipotesi tacita che vi debba essere un qualche livello astratto (ma quasi sempre non specificato) di descrizione condiviso da tutti i cervelli umani» (Hofstadter, 1995d, p. 515), che in realtà costituiscono un insieme eterogeneo essendo ogni cervello particolare diverso da ogni altro per quanto riguarda in special modo l’apparato di connessioni fra le cellule. In questo modo parlare di “cervello” equivale a utilizzare un termine connotato in maniera fortemente teorica all’interno dei sistemi di principi e proposizioni descrittive, le teorie, che tentano di spiegarne la natura. Non del tutto inverosimilmente si può affermare, dunque, che proprio la differenza fra il “cervello” astratto e universale e i “cervelli” particolari può essere considerata la prima forma di astrazione teorica verso una visione funzionale del sistema mente-cervello. Dopo tutto, dire che “il cervello umano può pensare” è come affermare qualcosa del tipo: «Esistono meccanismi astratti universali, che si realizzano in modo differente in ogni cervello specifico e che permettono che il pensiero abbia luogo» (*ivi*, p. 516). Tali meccanismi “cerebrali” «non sono *solo* componenti fisiche; piuttosto, sono strutture che si pongono in qualche punto dello spettro tra componenti fisiche e componenti immateriali, cioè tra hardware e software» (*ibidem*).

Tutto ciò porta a una visione unificata del sistema mente-cervello, in cui la spiegazione funzionale e l’individuazione delle controparti fisiche degli apparati funzionali non vanno viste in relazione di contrapposizione, bensì di giustapposizione, nel senso che, per un verso, a livello teorico è possibile dare, almeno in linea di principio, una descrizione fisica di ogni processo funzionale del sistema sia che si tratti, ad esempio, del meccanismo che regola la memoria a breve termine, sia che si prenda in considerazione l’unità basilare del cervello, il neurone; per un altro, appare chiaro che un’interpretazione dei meccanismi cerebrali non può non essere data in termini funzionali, sia che le funzioni espletate siano quelle dei singoli neuroni, sia che siano quelle di porzioni cerebrali come la parte del lobo frontale che regola la memoria a breve termine. L’interpretazione in termini di meccanismi in merito al sistema mente-cervello può essere definita, dunque, *pan-funzionale*, così come, d’altra parte, è richiesto dall’assunto che rende possibile la metodologia simulativa *tout court*, con tutte le conseguenze sulla sostituibilità dei supporti utilizzati, la quale diventa più una questione pratica, anche se d’innegabile importanza nel complesso della ricerca, di implementazione del livello scelto.

Come valutare, dunque, i modelli che si basano sull’impiego di microprocedure dopo questo *excursus*? Ovvero, queste costituiscono il livello di analisi migliore? In un’interpretazione “pan-funzionalista” la descrizione delle interazioni causali del meccanismo che realizza una funzione appare non scindibile dalla prestazione che realizza quella funzione, la quale ne costituisce l’altro aspetto esplicativo fondamentale. Infatti, se la nozione di “funzione” implica un meccanismo che compie la funzione, il risultato finale, esterno, processuale del meccanismo coincide con la funzione stessa. In questo modo, *anche* la valutazione della prestazione, cioè del modo in cui qualcosa (un meccanismo) *funziona* appare necessario per definirne la reale portata esplicativa, così come

prefigurato da Turing con le sue idee in merito alle macchine in grado di pensare. Se, dunque, estendiamo i criteri di Turing per valutare in via sperimentale ogni dispositivo meccanico, compreso il cervello umano, oltre i limiti tracciati dal gioco dell'imitazione, cioè oltre l'interazione in linguaggio naturale, si può ottenere un Test di Turing generalizzato che tenga in considerazione ogni prestazione compiuta da un sistema, la cui descrizione in termini meccanico-procedurali (o, se si vuole, in termini di procedura effettiva) è disponibile, senza tuttavia vincolare il meccanismo ai requisiti di localismo, finitezza e discretezza della computazione classica.

In questo modo, ovvero se la valutazione della effettiva simulazione di una funzione cognitiva è possibile sulla scorta della considerazione dell'insieme di meccanismo procedurale *e* prestazione, si ritorna alla questione del giusto livello di simulazione, che, fatto salvo come si è detto un elemento *convenzionale* imprescindibile in merito al *giusto livello*, è individuato dalle prestazioni compiute: «un modo non ambiguo di definire i livelli dei meccanismi è nei termini delle prove in grado di svelarli» (French, 1995, p. 146). Tali prove sono appunto le prestazioni messe in atto dal programma. Così Hofstadter (1995c, pp. 390-391) individua i criteri per la convalida dei modelli subcognitivi nella valutazione delle loro prestazioni, sottolineando che essi sono «qualitativi – di certo non quantitativi». La lista che viene fornita riguarda la plausibilità delle risposte, l'allineamento con le risposte umane, e ancora l'ovvietà, la verosimiglianza, l'eleganza, la creatività delle soluzioni date. Di questi criteri si può dire che sono *antropocentrici*, ma, in realtà, lo è anche la definizione di intelligenza che viene indagata dalle scienze cognitive, o almeno dall'IA fin dai tempi di Turing e Simon. Ancora, essi sembrano fortemente intrisi di *senso comune*, ma di fatto è proprio il senso comune che rende possibile giocare il gioco dell'imitazione. Infine, essi sembrano *intuitivi*, visto che possono essere ricavati «discutendo in maniera informale, con pochi interlocutori, senza bisogno di sperimentazioni psicologiche estese» (*ibidem*).

Se il discorso finisse qui, saremmo in presenza di un mero riproporre i criteri forniti da Turing con il suo gioco dell'imitazione, con l'unica eccezione che quelli proposti da Hofstadter virano oltre la rotta stabilita dal paradigma dell'interazione in linguaggio naturale e si allargano fino a comprendere altri aspetti di una *sensatezza* intuitiva forse troppo *naïve* e lontana da criteri di controllabilità oggettiva. D'altra parte, la nozione di "intelligenza" definisce un concetto limite, un'idea regolativa dell'IA e il fatto che sia l'intelligenza (umana) a riconoscere intuitivamente l'intelligenza di un sistema (umano o artificiale) fa parte dell'impulso filosofico ed epistemologico connaturato alla ricerca in IA e nelle scienze cognitive. Tuttavia, occorre considerare ancora una volta che il livello che si è scelto di simulare è quello delle microprocedure, fatto non privo di conseguenze.

Innanzitutto, è proprio French (1990) a considerare la possibilità di una revisione del Test di Turing, inteso come un'estensione del gioco dell'imitazione all'effettiva valutazione di un sistema artificiale in grado di pensare, e non soltanto come criterio filosofico di definizione della nozione di "intelligenza". Lo scoglio contro cui cozzerebbe, secondo French, un qualsiasi programma che

affrontasse il Test di Turing sarebbe proprio quello relativo a domande che interessano compiti subcognitivi, cioè che riguardano i processi di categorizzazione, nel senso di apprendimento per somiglianze da un insieme di input esterni, e costruzione di analogie. Questo avverrebbe perché «le risposte ai quesiti subcognitivi emergono da una lunga esperienza di vita con i dettagli dell'esistenza, che va da una conoscenza del mondo funzionalmente adattiva a quella delle inutili banalità del quotidiano» (*ivi*, p. 63). In tali parole possono essere ravvisate due punti fondamentali. Il primo è la centralità ancora una volta riconosciuta ai processi subcognitivi, espletati nel modelli dalla compagine globale delle microprocedure, non soltanto come livello adeguato cui indagare l'"intelligenza", ma anche come direttamente implicati, loro e non altri, nei processi di categorizzazione e concettualizzazione. Il secondo riguarda il fatto che, se la revisione del Test di Turing è possibile, essa non deve implicare l'abbandono della "prova da prestazione", quanto piuttosto cercare di scovare nuovi modi di guardare alle prestazioni del programma, oltre l'interazione in linguaggio naturale. E questo, cosa che può sembrare ovvia, proprio perché le capacità cognitive che si stanno indagando sono solo indirettamente legate con la capacità di interagire in linguaggio naturale. Linguaggio e intelligenza si devono in qualche modo *scollare* per far sì che ci possa essere una corretta valutazione dei prodotti della ricerca simulativa. Forse in questo può essere vista consistere la "mossa realistica" di cui si diceva nel primo capitolo, a chiusura del cerchio che porta le discipline interessate a testare i manufatti artificiali in grado di produrre intelligenza dapprima come manipolatori di simboli vuoti, e infine come sistemi in grado di padroneggiare l'esperienza, segmentandola, categorizzandola, producendone rappresentazioni *epistemicamente significative* dal punto di vista del sistema stesso.

L'idea di una convalida dei modelli subcognitivi attraverso l'analisi delle prestazioni si rispecchia nei tre criteri aggiuntivi proposti a questo scopo da Hofstadter (1995c, p. 391). Essi riguardano non soltanto il risultato, bensì i processi compiuti dal programma. I primi due si riferiscono: 1) alla *plausibilità* del processo *per un osservatore esterno*; 2) *all'allineamento del programma alle prestazioni degli esseri umani* una volta che si siano modificate le componenti strutturali dell'architettura, in particolare la composizione della rete semantica che esprime la conoscenza del programma, per riflettere il cambiamento del contesto in cui gli esseri umani vengono fatti operare. Si pensi, ad esempio, al caso in cui ad un soggetto venga detto di trovare soluzioni che sfruttino in modo particolare relazioni di simmetria fra gli elementi del dominio, o alcuni elementi specifici del dominio, e così via.

Non si può negare che il primo metodo richiama in parte quello utilizzato da Newell e Simon in tempi più remoti, basato sull'utilizzo di protocolli introspettivi. Il secondo criterio costituisce, dunque, una sorta di controprova del precedente e condivide alcuni aspetti con il *priming effect*, nella misura in cui sia il programma che l'agente umano subiscono analoghi condizionamenti nelle strategie di ricerca della soluzione. Tuttavia, l'aspetto vantaggioso di questi criteri sta nel fatto che il loro utilizzo sposta l'attenzione da ciò che il programma produce a *come* lo produce. In

quest'ottica, inoltre, acquisiscono senso i numerosi esperimenti su soggetti umani compiuti parallelamente allo sviluppo di quasi ogni modello, di cui si trova ampia trattazione, ad esempio, in French (1995), McGraw (1995), Rheling (2001), Foundalis (2006). Al di là delle critiche di "povertà esplicativa" espresse da Hofstadter nei confronti della metodologia statistica della psicologia sperimentale, il raffronto con le prestazioni degli esseri umani assume un ruolo centrale nella valutazione comparata delle prestazioni di individui e modelli, non perché gli uni *o* gli altri vengono sottoposti a determinati esperimenti, ma perché gli uni *e* gli altri vengono sottoposti agli stessi esperimenti, instaurando un legame di *covarianza funzionale* che trasforma gli effetti contestuali prodotti sugli uomini in cambiamenti strutturali nei modelli.

Infine, l'ultimo criterio proposto da Hofstadter (1995c, p. 391-392) si può considerare *totalmente funzionale*, poiché implica l'elisione di alcune parti dell'architettura al fine di sperimentare il loro effettivo contributo al processo di ricerca descritto dal modello e implementato nel programma. L'elisione di parti dell'architettura richiama analoghi metodologie di *testing* utilizzate nei sistemi connessionisti e, più in generale e in maniera speculare, le metodologie largamente impiegate dalla neuropsicologia per l'individuazione dei compiti cui sono preposte specifiche aree cerebrali. L'impiego di questo metodo, in primo luogo, mostra una certa robustezza nei modelli, caratteristica che viene spesso menzionata fra i pregi delle reti neurali, pur non essendo tali modelli definibili come connessionisti. Tuttavia, il fatto che essi possano essere elisi perdendo gradualmente le loro capacità indica una loro eterogeneità rispetto ai modelli tradizionali, maggiormente caratterizzati dai vincoli della computazione classica, in particolar modo dalla mono-serialità delle operazioni. Va, comunque, sottolineato che questo è un aspetto condiviso in tutte quelle architetture complesse di tipo eterarchico, in cui l'elaborazione non deve compiere un unico percorso obbligato e i vari moduli arricchiscono le possibilità della strategia di ricerca. Da questo punto di vista, i modelli subcognitivi possono essere considerati analoghi a questo tipo di architetture. Inoltre, va aggiunto che un vero e proprio blocco nel programma in mancanza anche soltanto di un piccolo passaggio funzionale, cioè un'operazione comandata da un'istruzione, è un discorso che riguarda più gli aspetti implementativi dell'algoritmo che non il modello stesso. Tuttavia, le possibilità di analisi che le lesioni dei modelli riservano sono molte e in misura proporzionale alla quantità degli elementi in gioco *simultaneamente* (dal punto di vista del modello e non della sua implementazione), e dunque delle microprocedure attive.

Esperimenti di questo tipo sono stati tentati con COPYCAT e la Mitchell vi dedica un intero capitolo (1993, pp. 183-199). Tuttavia, le loro conseguenze sono facilmente estendibili anche agli altri modelli. In particolare, tali esperimenti riguardano sia la rete concettuale, attraverso il livellamento della profondità concettuale o la fissazione delle lunghezze dei collegamenti della rete, che porta a una minore elasticità in termini di influenze contestuali *top down* dovuta alla minore capacità di slittamento fra i concetti; sia il vincolo della variabile temperatura a differenti valori, con il quale si arriva a determinare un'elaborazione molto o poco basata sulla casualità (rispettivamente

se il valore viene fissato al massimo o al minimo), o in generale a una scarsa flessibilità nella ricerca di visioni alternative una volta costituita una prima visione globale della situazione (se il valore viene fissato a un livello intermedio); sia, infine, la soppressione di determinati insiemi di microprocedure, che porta o alla incapacità di procedere alla formazione di determinate strutture percettive o, nel caso di una diminuzione generalizzata delle microprocedure, a una drastica riduzione del processo di ricerca parallelo, sempre con grave danno sulle complessive capacità esplorative della situazione da parte del programma.

Si può concludere che questo ultimo criterio di valutazione dei modelli costituisce una sorta di *experimentum crucis* riguardante le varie funzioni modellate dall'architettura nel suo complesso. Ciò che i risultati di questo metodo evidenziano è, soprattutto, il ruolo innegabile rivestito dal *parallelismo*, attuato nei modelli dall'impiego delle microprocedure, circostanza che porta direttamente alla considerazione di quale tipo di computazione mettono in atto i modelli subcognitivi e, conseguentemente, ad ulteriore approfondimento della posizione da essi occupata all'interno delle discipline simulative.

4.4 Microprocedure e computazione: il paradigma della creatività

Dal lato implementativo le microprocedure, si è visto, corrispondono a semplici operazioni che il programma può compiere. Se si guarda, però, ai modelli dal punto di vista della loro architettura globale, essi possono facilmente essere descritti come architetture modulari complesse, alla stregua di SOAR o ACT-R per citare solo le più famose[3], pur con caratteristiche peculiari loro proprie.

Lo schema generale dei modelli subcognitivi, delineato nel secondo capitolo è costituito da una doppia struttura rappresentativa, che differenzia, accorpandoli distintamente nell'architettura del programma, due sensi di "rappresentazione della conoscenza". Nella memoria a lungo termine è rappresentata buona parte della conoscenza permanente del programma, che esso è in grado di utilizzare in maniera dinamico-adattiva. Nella memoria a breve termine è rappresentata la conoscenza che il programma si fa della situazione percepita. Si tratta, perciò, di una rappresentazione *dipendente dall'esecuzione*, mentre la conoscenza concettuale è, almeno in parte, *indipendente dall'esecuzione* del programma. Le varie espansioni dei modelli non modificano questa impostazione di fondo, ma, come abbiamo visto, aumentano le possibilità del modello di muoversi in domini sempre più ricchi e vicini ad ambiti del mondo reale, anche se ciò non va confuso con l'idea, dal sapore antico, che questi modelli possano essere sulla giusta strada per sviluppare una simulazione dell'intelligenza *tout court*, comprensiva di tutti gli aspetti cognitivi ed emotivi che la caratterizzano. D'altra parte, l'intento esplicito degli autori dei modelli che abbiamo

[3] Per una presentazione di SOAR si rimanda a Laird, Newell, Rosenbloom (1987). Per ACT-R si veda Anderson, Lebière (1999).

considerato è quello di simulare soltanto alcune ben determinate capacità del pensiero, che coinvolgono meccanismi di rappresentazione *così come* strutture rappresentative già formate.

Per realizzare tale intento è necessaria una componente procedurale, la quale, seppur in parte etero-guidata, rappresenta altra conoscenza permanente del programma, una conoscenza però operativa, un *sapere come* piuttosto che un *sapere che*, per riprendere la nota distinzione suggerita da Gilbert Ryle (1949). Una conoscenza di questo tipo è rappresentata, appunto, *nelle* microprocedure. Dal punto di vista della *computer science* le microprocedure sono, perciò, l'implementazione di una struttura interattiva complessa basata sullo scambio di informazione fra agenti autonomi. Abbiamo ampiamente descritto il genere di operazioni che tali agenti sono chiamati a compiere. Essi, parlando generalmente, si caratterizzano per la semplicità d'azione di contro alle prestazioni effettuate dal programma *sia per quanto riguarda il risultato, sia in riferimento alle macroazioni compiute* di cui si trova traccia in appositi moduli dedicati o nell'analisi a posteriori dell'esecuzione, ovvero comunque ad un *meta-livello rispetto a quello dell'elaborazione microprocedurale*.

La convalida dei modelli attraverso il metodo delle lesioni, oltre a indicare la loro composizione modulare, ci fornisce un'indicazione sul modo in cui va intesa la computazione attraverso microprocedure, altrimenti detta "multi-agente". Infatti, le lesioni producono effetti massivi, così che le microprocedure possono essere considerate compiere funzioni modulari globali, definibili, ad esempio, come "formazione di gruppi", "atteggiamento esplorativo ampio", "creazione di corrispondenze", "richiamo (per attivazione) di concetti astratti", e così via. A secondo della particolare funzione cognitiva che si vuole sperimentare è possibile variare il numero delle microprocedure che la espletano, aumentandole o diminuendole fino ad eliminarle del tutto; ciò è effettuabile anche in maniera indiretta modificando dall'esterno la variabile temperatura o l'apporto *top down* dei concetti che si traduce nell'immissione di determinate microprocedure.

Questo tipo di esperimenti non è stato tentato in tutti i modelli, anche se appare essere uno strumento investigativo molto potente. Da una parte, infatti, rende possibile una gradualità nel modo di compiere una funzione da parte del programma, il che garantisce stime più esatte dell'efficacia della funzione stessa di quelle concesse da una sua alternanza binaria, tutto-o-niente, di presenza e assenza; dall'altra, permette di collegare fenomeni cognitivi globali ai loro costituenti, i *parametri cognitivi* visti in precedenza, che compiono effettivamente il processo. Qui entrano in gioco le nozioni di "emergenza" e di "comportamento emergente", le quali stanno a significare che un processo a un certo livello è il risultato del complesso delle azioni di processi al livello inferiore. La concreta realizzabilità di queste nozioni nelle architetture complesse basate su computazioni multi-agente dovrebbe far riflettere sul fatto che l'idea secondo cui "il tutto è superiore alla somma delle parti" trova un riscontro effettivo, reale si vorrebbe dire, se applicato ad una ambito processuale, piuttosto che a quello linguistico-simbolico, per il quale si traduce, d'altro canto, in una visione olistica del significato. Tuttavia, non è questa la sede per affrontare i pro e i contro di tale punto di

vista. Vale, però, la pena sottolineare almeno la stretta correlazione fra olismo ed emergentismo, anche se sono concetti teorici che differiscono, generalmente, quanto a universo di riferimento e di applicazione.

Dunque, quale conclusione si può trarre sul tipo di computazione che caratterizza i modelli subcognitivi? Essa è classica o ibrida o connessionista? Si può considerare dinamica? O, anche, un genere di computazione che estende i limiti della Turing-computabilità? Come è noto, dietro a ognuna di queste etichette è stato fatto rientrare un paradigma ontologico-metodologico relativo allo studio dei fenomeni cognitivi. L'analisi per paradigma rischia, tuttavia, di fuorviare la comprensione dei modelli subcognitivi e per rispondere alle domande che ci siamo posti vanno enucleati gli aspetti salienti del processo computazionale in atto.

La questione in merito al tipo di computazione attuato dai modelli subcognitivi sembra possa avere una risposta non univoca. Come abbiamo visto, nessuno dei modelli presi in considerazione, ad eccezione di uno solo, LETTER SPIRIT 2 (Rehling, 2001) si avvale di reti neurali per la rappresentazione e l'elaborazione della conoscenza, e anche in quell'unico caso, si tratta di un'applicazione ad una parte soltanto della conoscenza rappresentata nel programma. Non si può, dunque, considerarli modelli connessionisti, almeno nel senso tipico del termine. Tuttavia, il fatto che essi, pur utilizzando una rappresentazione della conoscenza in forma simbolica, non trattino questa informazione attraverso una manipolazione logico-sintattica, sembra negare la possibilità di una loro definizione come sistemi di IA simbolica classica. Le seguenti affermazioni della Mitchell e di Hofstadter in merito a COPYCAT sembrano confortare questa duplice esclusione paradigmatica:

> COPYCAT è un programma computazionale progettato per essere in grado di scoprire analogie penetranti in modo realistico, dal punto di vista psicologico. *La sua architettura non è simbolica, né connessionista, né un ibrido tra le due* (benché alcuni potrebbero considerarla tale); il programma, piuttosto ha un tipo nuovo di architettura che si situa fra i due estremi. Essa è *emergente*. (Mitchell, Hofstadter, 1994, p. 225 [enfasi mia])

Questa parole si riferiscono a COPYCAT ma possono essere considerate valide per tutti i modelli dello stesso tipo. Devono, tuttavia, essere precisate. Il fatto che tali modelli non sono visti come ibridi sta qui ad indicare soltanto che essi non si avvalgono di differenti moduli dedicati in relazione tra loro e implementati simultaneamente, alcuni simbolici e altri connessionisti. D'altra parte, la definizione di sistema ibrido è piuttosto basata sulla capacità di sfruttare entrambe le potenzialità della dicotomia discreto/continuo, ovvero i sistemi ibridi «sono caratterizzati dal processare dinamicamente informazione simbolica e dall'interpolare almeno un processo nel quale variabili continue ricorrono in maniera essenziale fra due processi discreti» (Sandri, 2006, p. 210). Tali

sarebbero ad esempio i sistemi analogici[4]. I modelli subcognitivi potrebbero essere fatti rientrare sotto la definizione di sistemi ibridi proprio in virtù del fatto che la loro elaborazione è frutto dell'interazione di numerosi processi di basso livello, le microprocedure, i quali determinano un andamento continuo nel tempo dell'elaborazione influenzando le variazioni della rete semantica e della variabile temperatura. Quest'ultima, seppure dal punto di vista implementativo possa assumere soltanto valori compresi in un *range* predefinito (in genere da 0 a 100), è dipendente da numerosi processi integrati e non deterministici espletati dalle microprocedure, i quali dipendono a loro volta sia direttamente dagli elementi percepiti nello spazio di lavoro, sia retroattivamente dall'influenza generata dalla parte semantica del programma e dalle variazioni della temperatura. Per tali ragioni, sembra appropriata la definizione di questi modelli come ibridi data da Kokinov e French (Kokinov, French, 2003). Tuttavia, la loro affermazione in merito al fatto che i modelli subcognitivi sono una «combinazione sia dell'approccio simbolico che di quello connessionista» (*ivi*, p. 115) è corretta solo se la combinazione non si riferisce all'impiego *in toto* di entrambe le metodologie simulative, ma di alcuni aspetti di esse. Come abbiamo visto, infatti, la dimensione simbolica dei programma è molto accentuata e ne costituisce un tratto fondamentale, ma i modelli mettono in atto un processo emergente proprio grazie a un andamento parallelistico e a una rappresentazione della conoscenza basata su connessioni ricorrenti, variabili in modo continuo.

Naturalmente, essendo questi modelli implementati su macchine rigide, seriali e sequenziali, il processo di elaborazione può essere arrestato e ripreso in qualsiasia momento senza subire variazioni di sorta. Ciò sembrerebbe far venir meno la caratteristica di dinamicità dei sistemi. Questo ha, però, tutta l'apparenza di un falso problema. Infatti, solo nel caso in cui il sistema interagisse con l'ambiente esterno, esso subirebbe una modificazione indotta da una situazione di arresto e ripresa dell'elaborazione. Questo non accade nei modelli subcognitivi, la cui informazione è già codificata in forma virtuale e statica in una delle memorie del modello. Tale situazione appare in linea con il rispetto dei vincoli della computazione classica, perlomeno quanto alla chiusura col mondo esterno. Tuttavia, qui la computazione non è chiusa, perché non si può affermare questo in riferimento al fatto che gli elementi del dominio sono simulati e non, invece, tradotti dai canali interfaccia in dati maneggiabili dall'*hardware* su cui il sistema è implementato. L'aspetto principale rimane, dunque, il fatto che l'elaborazione è probabilistica e non può essere prevista all'inizio, cioè il sistema non è riducibile, *a priori*, a un automa a stati finiti deterministico.

I modelli subcognitivi si situano, dunque, oltre i limiti della Turing-computazione o della computazione classica (Sigelmann, 1999; Sandri, 2006)? La risposta sembra affermativa. Il titolo di un importante saggio hofstadteriano dei primi anni ottanta evidenzia chiaramente l'intento di

[4] In letteratura sono reperibili due definizioni di processo analogico su cui può basarsi un sistema. Processi analogici sono quelli che utilizzano variabili continue, ma anche, in un'altra interpretazione, il processo analogico è riferito alla relazione di somiglianza che si instaura fra la rappresentazione e il rappresentato, sia che si tratti di un sistema biologico (animale o umano), sia che ci si riferisca a un sistema artificiale. Su questo tema si rimanda a Cordeschi, Frixione (2006).

distaccarsi dall'IA tradizionale, proponendo come necessario un «risveglio dal sogno booleano» e l'adozione della «subcognizione come computazione» (Hofstadter, 1985f, p. 631). Diverse sono, inoltre, le metafore utilizzate per descrivere il processo elaborativo di questi modelli riprese dalla biologia: il metabolismo cellulare (Hofstadter, 1983a), il sistema immunitario (Mitchell, 2001), le colonie organizzate di insetti come le formiche (Mitchell, 2005). Tali metafore indicano che i modelli subcognitivi si ispirano fortemente all'*organizzazione funzionale* che permette l'espletamento dei compiti da parte dei vari sistemi biologici. Il tipo di organizzazione funzionale che essi mettono in atto è una strategia parallelistica che evolve in base all'informazione messa a disposizione dall'ambiente attraverso una dinamica adattiva, la quale permette di costruire la rappresentazione della situazione in modo da includere progressivamente in una visione coerente tutti gli elementi dello spazio percettivo e di operare di conseguenza per arrivare a una soluzione. È abbastanza intuitivo, ad esempio, il parallelo con il sistema immunitario, che adatta la produzione di anticorpi agli antigeni presenti nell'organismo attraverso un sistema di innumerevoli microazioni esplorative-costruttive in maniera molto simile ad una strategia di prova ed errore.

I modelli subcognitivi possono, dunque, a ragione considerarsi *sistemi complessi adattivi* (Mitchell, 2001), dei quali può darsi un'interpretazione in termini di *dinamicismo*, nella misura in cui il processo si svolge in modo continuo fra pressioni semantiche *top down* e stocastiche *bottom up* (French, in corso di pubblicazione), cioè attraverso un andamento ibrido vincolato a un doppio contesto (Kokinov, French, 2003). Poiché si tratta di sistemi basati su metodi stocastici per la risoluzione di problemi, possono essere fatti rientrare nel filone della *Natural Computation*[5], nel quale, come fa notare Sandri, l'attenzione viene spostata dalla computazione di funzioni alla computazione come processo di elaborazione e trasmissione dell'informazione secondo dinamiche continue:

> Gli aspetti di comunicazione di [tali] sistemi di computazione sarebbero non descrivibili entro il metodo classico della computazione, in quanto si tratterebbe di comunicazione fra componenti computazionali entro il sistema e di comunicazione fra il sistema computazionale e l'ambiente: e queste proprietà sarebbero in conflitto con le proprietà di sistema chiuso e completo della computazione classica. Nel sistema di computazione interattiva (interazione con altre componenti computazionali e con l'ambiente), la comunicazione interverrebbe durante la computazione, mentre il sistema chiuso processa un input dato all'inizio della computazione. (Sandri, 2006, p. 218)

Queste parole possono ben descrivere i modelli subcognitivi. La conclusione cui arriva Sandri è che «i processi interattivi non sarebbero algoritmici: entro questi processi gli inputs sono influenzati dagli outputs, e la proprietà renderebbe non funzionale in senso stretto il processo computazionale» (*ibidem*). In effetti, nei sistemi subcognitivi algoritmiche in senso stretto sono le microprocedure. Di

[5] Si veda, ad esempio, Eiben, Rudolph (1999).

conseguenza, è la loro interazione, corredata di scambio informazionale, a rappresentare un'uscita dal computazionalismo classico, anche se i processi emergenti di livello superiore possono ancora essere interpretati in senso funzionale, seppure, proprio per le ragioni viste prima nelle sperimentazioni con i processi di lesione, in un'accezione indebolita di "funzionale" rispetto alla quale una determinata funzione è svolta in maniera robusta, ridondante e flessibile da un insieme di agenti operativi specializzati nel realizzare, o declinare, in modi diversi la stessa funzione principale (si pensi alle diverse microprocedure che realizzano la medesima funzione di creazione di corrispondenze o di gruppi).

Se i processi interattivi e dinamici «propongono la costruzione di nuovi paradigmi» (*ibidem*), è anche vero che c'è stata un'ampia proliferazione negli ultimi anni di paradigmi computazionali differenti, ciascuno legato allo sfruttamento di diverse euristiche di computazione. Si considerino, ad esempio, e solo per rimanere nell'ambito della *Natural Computation*, gli Automi Cellulari, gli Algoritmi Evolutivi, gli Algoritmi Genetici, la *DNA Computing*. Poiché l'impostazione dei modelli subcognitivi costituisce in ogni caso un approccio a sé stante all'interno delle scienze cognitive è forse ozioso ricercare quale di questi paradigmi possa essere il più vicino a quello subcognitivo. Proponiamo pertanto, se proprio si vuole inserire in un paradigma tale approccio, di includerlo in un generale "paradigma della creatività", che include differenti tipi di computazione algoritmica.

Lo studio dei processi creativi è stato ampiamente affrontato all'interno delle scienze cognitive e dell'IA[6]. È Hofstadter stesso a parlare del tentativo «apparentemente paradossale di meccanizzare la creatività» (Hofstadter, 1985b). A distanza di più di venti anni da quella proposta è chiaro come essa si sia concretizzata proprio grazie a forme di computazione interattiva, dinamica, stocastica, in cui il processo e la trasmissione di informazione all'interno di una sistema costituiscono gli aspetti essenziali per la produzione di rappresentazioni (via l'informazione trasmessa), le quali permettono la realizzazione di prestazioni creative, come si è visto nel corso del capitolo precedente. Quali sono le caratteristiche principali che autorizzano a parlare di processi creativi all'interno dell'approccio subcognitivo?

Nella ricostruzione dell'evoluzione dei modelli subcognitivi abbiamo constatato come un'attenzione sempre maggiore alla creatività scaturisca dall'impostare modelli il cui dominio di applicazione è sempre più ricco ed articolato. In particolare, negli studi compiuti in merito alla produzione di stili alfabetici in LETTER SPIRIT e in LETTER SPIRIT 2 sia McGraw (1995) sia Rehling (2001) affrontano questo problema in maniera approfondita. McGraw soprattutto sottolinea che la creatività nei modelli subcognitivi è strettamente legata alla nozione di "casualità" e all'impiego di un metodo di ricerca, come la scansione parallela a schiera, che si avvale di processi casuali supervisionati da appositi dispositivi di controllo. Ed è proprio grazie al sapiente equilibrio fra *flessibilità* e *controllo* che questi modelli possono dirsi compiere *processi creativi autonomi* (McGraw, 1995, pp. 111 e sgg.). Infatti, una misura troppo elevata di casualità indebolisce il

[6] Si vedano Boden (1990, 1994), Johnson-Laird (1993) e Dartnall (2002).

processo di ricerca rendendolo troppo dispersivo, ma un controllo eccessivo da parte di un supervisore esterno, come può essere il programmatore, rende il processo creativo non autonomo di fatto dissolvendolo. La corretta integrazione di questi accorgimenti, attraverso dispositivi di meta-controllo che variano dinamicamente nel tempo, sia per quanto riguarda l'auto-valutazione del programma espressa dalla temperatura, sia in merito alle variazioni nella componente epistemica costituita dalla rete concettuale, permette il dispiegarsi di quel "ciclo centrale retroattivo della creatività" che corrisponde alla struttura basilare della TCCL menzionata nel secondo capitolo. Creatività e circolarità rappresentativa, attraverso il passaggio di informazione fra le varie componenti del programma, sono due aspetti complementari del processo di elaborazione dei modelli subcognitivi, direttamente co-implicati nell'esibizione di un comportamento *emergente*.

L'elaborazione emergente si configura, dunque, come processo fortemente creativo, in grado di dominare le interferenze nell'elaborazione dovute all'apporto di informazione dall'esterno, circostanza che porta a considerare questi modelli pienamente in linea con i principi della cognizione situata. Tuttavia, il fatto che si possa parlare di elaborazione emergente è ancora una volta strettamente determinato dall'impiego di processi interattivi fra microprocedure e dall'utilizzo di strategie per il controllo dell'andamento stocastico dell'elaborazione. Si può pertanto ricondurre la questione della creatività ai quattro principi proposti dalla Mitchell (2005) per ogni sistema "decentralizzato" sia artificiale che biologico (ad esempio, il già ricordato sistema immunitario), che voglia dirsi dotato di opportune funzioni di controllo e auto-consapevolezza e allo stesso tempo esibire un comportamento non deterministico:

- l'informazione globale deve essere codificata in forma di schemi (*patterns*) statistici e dinamici attraverso le componenti del sistema;
- la casualità (*randomness*) e la probabilità sono essenziali;
- il sistema deve eseguire una ricerca a grana fine e parallela delle possibilità;
- il sistema deve esibire una continua interazione di processi *top down* e *bottom up*.

Come si vede, la funzione di controllo e supervisione viene espletata dall'interazione fra processi percettivi e cognitivi, collegati alle diverse forme di memoria di un sistema cognitivo contenenti differenti tipi di informazione. La portata della casualità viene arginata anche dall'utilizzo di funzioni o algoritmi probabilistici, vincolati a punti convergenza o attrazione o stabilizzazione da processi dinamici di inclusione della nuova informazione esperita all'interno di schemi precostituiti, così come, ad esempio, in una colonia di formiche è il segnale (chimico) lanciato da un esploratore che ha trovato *casualmente* il cibo a rinforzare la possibilità che altre formiche percorrano la stessa strada in base all'imperativo istintuale (assimilabile a una conoscenza innata) del cercare cibo per il nutrimento. Un sistema biologico che non avesse questo tipo di conoscenza (imperativa) non solo perirebbe velocemente, ma gli mancherebbero *a priori* le condizioni per svilupparsi.

In conclusione, le microprocedure rivestono un ruolo centrale dal punto di vista della computazione emergente, permettendo lo svolgimento di funzioni in maniera robusta grazie alla loro *ridondanza operativa*. Rimane da vedere, per ottenere un'ulteriore verifica di plausibilità dei modelli cognitivi artificiali basati su questo tipo di elaborazione, se è possibile trovare un loro corrispettivo dal punto di vista cerebrale.

4.5 Microprocedure e cervello: la teoria dei simboli attivi

È lecito chiedersi fino a che punto è possibile aspettarsi di trovare un correlato neurale delle microprocedure, cioè dei processi subcognitivi, che, ricordiamolo, sono caratterizzati proprio dal ricadere sotto la soglia dell'attenzione cosciente. Anche se la ricerca di un tale correlato non rientra negli intenti di coloro che hanno sviluppato i modelli simulativi subcognitivi, va compreso almeno a quali condizioni sarebbe possibile individuare la realizzazione cerebrale di tali microprocedure, individuate in maniera funzionale, e, soprattutto, come arrivare a una conferma sperimentale di questo fatto.

Per quanto riguarda il secondo aspetto, la conferma sperimentale incorre nei problemi che abbiamo esposto all'inizio di questo capitolo in merito alla valutazione della ricerca simulativa rispetto al possibile raffronto con i risultati delle neuroscienze. Ad esempio, recentemente è stato approntato un esperimento nel quale alcuni soggetti sottoposti a risonanza magnetica funzionale sono stati invitati a risolvere i problemi di analogia affrontati da COPYCAT e METACAT (Geake, Hansen, 2005). I risultati di questo esperimento hanno dimostrato che nello svolgere questo compito i soggetti utilizzano una determinata area della corteccia prefrontale che può essere considerata causalmente anteriore a, e dunque implicare nelle sue conseguenze l'attività di, altre aree cerebrali dedicate ad attività cognitive superiori come la formazione di regole e l'associazione a distanza, nonché al prendere decisioni. Tuttavia, l'interpretazione dei risultati va sempre presa con le dovute cautele, proprio perché l'analisi è stata condotta su una prestazione complessiva e non sul comportamento di singole microprocedure, per le quali non è neppure stato ipotizzato un possibile corrispondente neuronale.

In che modo, dunque, è possibile ipotizzare una realizzazione neuronale per esse? La risposta non può che essere ipotetica e deve tenere conto di che cosa sono effettivamente le microprocedure e del motivo per cui vengono introdotte. Esse sono, infatti, agenti operativi, i quali però vengono individuati in senso funzionale per rendere possibile l'implementazione di sistemi artificiali simulativi che non rinuncino a sviluppare un potere rappresentazionale e che, dunque, siano basati sulle potenzialità del simbolico, anche se svincolato dalle rigide procedure di manipolazione sintattico-formali esposte alle obiezioni di Searle, sintetizzate nel primo capitolo, ormai considerate prototipiche. Tuttavia, costruire un sistema di rappresentazione autonomo e non formalmente vuoto

ha il costo (teorico) di esporre il sistema al rischio della *regressio ad infinitum* della mente come "teatro cartesiano" (Dennett, 1998).

La scomposizione della funzione rappresentazionale in microprocedure salva da questo pericolo, se si accetta come valida la nozione di elaborazione emergente. Questa linea teorica è fortemente condivisa da Dennett attraverso la sua teoria degli *homunculi* (1978, 1991) e riproposta ancora recentemente in Dennett (2005) dove si legge: «finché i vostri *homunculi* saranno più stupidi e ignoranti dell'agente intelligente che compongono, l'operazione di nidificare *homuniculi* all'interno di *homunculi* può arrivare ad un punto finale, raggiungendo il livello più basso in cui vi sono agenti così modesti da poter essere rimpiazzati da macchine» (*ivi*, p. 131). Va detto che la teoria formulata da Dennett è una teoria filosofica sulla mente, o sul sistema mente-cervello. L'affinità con le idee alla base dei modelli subcognitivi a questo punto dovrebbe, però, essere manifesta ed è in più giustificata dalle parole espresse dallo stesso Dennett nella prefazione al libro di French (1995) in favore dell'approccio da lui adottato. Secondo Dennett, infatti, French «modella un fenomeno che non è difficoltoso né assolutamente invisibile ma piuttosto *appena* fuori della possibilità di raggiungimento per l'introspezione del lettore» Questi «eventi quasi-introspettivi» vanno considerati accadere «immediatamente dietro le quinte» del teatro cartesiano «sul palcoscenico del quale sfila la parata della coscienza» (*ivi*, p. viii).

Queste parole evidenziano, tuttavia, che la teoria di Dennett è sorta anche nel tentativo di fornire una possibile spiegazione ad alcuni fenomeni mentali relativi alla coscienza, ai *qualia*, al soggettivismo, unitamente alla proposta di un approccio "eterofenomenologico" alla questione[7]. Manca un'effettiva specificazione delle microprocedure, una difesa delle quali dal punto di vista computazione, peraltro, è implicita nel fatto che i modelli proposti funzionano senza andare in *loop* e senza cedere nel regresso all'infinito della rappresentazione, come fa notare Hofstadter (2001, p. 538). Perciò, se la controprova migliore della plausibilità della teoria sta nel funzionamento fattuale dei modelli, sembra che il modo migliore di definire la loro natura è considerare quale ruolo effettivo esse ricoprano.

In tale prospettiva, le microprocedure possono essere considerate alla stregua di *rivelatori* di proprietà e relazioni, o anche soltanto di proprietà, se vogliamo considerare le relazioni come proprietà che si riferiscono a più argomenti. Nel cervello andrebbero, dunque, cercate controparti a questi agenti funzionali procedurali del programma. Un'ipotesi plausibile è che esse possano essere rinvenute in forma di *pattern* neurali, strutture formate dai neuroni, dai loro collegamenti e dal potenziale di attivazione determinato chimicamente, quest'ultimo essendo l'unico vero aspetto sempre variabile dell'apparato neuronale. Dal punto di vista teorico esse si inseriscono senza

[7] Si vedano Dennett (1982, 1991) per la definizione di "eterofenomenologia". Ad esempio, essa è caratterizzabile come «un sentiero *neutrale* che ci conduce dalla scienza fisica oggettiva, e dalla sua insistenza sulla prospettiva in terza persona, a un metodo per la descrizione fenomenologica che può (in linea di principio) rendere giustizia delle esperienze soggettive più private e ineffabili pur senza mai abbandonare gli scrupoli metodologici della scienza» (Dennett, 1991, p. 86).

difficoltà in una visione che abbracci un'ontologia processuale, piuttosto che oggettuale, come quella proposta, ad esempio, da Manzotti e Tagliasco (2006) per spiegare i fenomeni mentali coscienti, in cui ancora una volta ciò che conta è la dinamicità del processo piuttosto che la sua cristallizzazione attorno ad attrattori che esprimono i massimi delle equazioni dinamiche che li descrivono.

Se, tuttavia, si vogliono lasciare da parte le questioni relative alla coscienza e ai *qualia*, meritevoli di una trattazione più approfondita che qui non può avere luogo, si ritorna al *problema della conoscenza* e al modo in cui viene risolto attraverso le microprocedure. Queste, in quanto rivelatori o recettori di proprietà basilari di corrispondenza e somiglianza, sia a livello di categorizzazione che di mappatura concettuale, sono concepite per andare a catturare quelle primitive percettive (*features*) grazie ad un'opera di filtraggio del materiale nel dominio, come si è visto soprattutto nei modelli subcognitivi più recenti. Se le teorie che richiamano questo tipo di processo, ad esempio quelle della Treisman e di Biederman, sono valide e costituiscono un punto di appoggio per questo approccio simulativo, va comunque precisato che esso estende, attraverso la formulazione di opportune architetture, la dinamica costruttivista applicandola anche alle strutture concettuali rappresentante nel sistema.

Le microprocedure che caratterizzano i modelli computazionali esaminati e permettono la loro applicazione in domini sempre più complessi mostrano, dunque, di essere strettamente legate alla rappresentazione concettuale della conoscenza dei sistemi. Per arrivare ad essa occorre passare attraverso l'antecedente delle microprocedure all'interno dell'impostazione subcognitiva: la teoria dei simboli attivi.

Questa teoria viene esposta da Hofstdater in *Gödel, Escher, Bach*. Egli propone di chiamare «*simboli* [i] complessi neuronici, o moduli neuronici, o pacchetti neuronici, o reti neuroniche, o unità multineuroniche» (Hofstadter, 1979, p. 378), che ipotizza corrispondere a ogni concetto. «I simboli sono le realizzazioni circuitali, quindi appartenenti allo hardware, dei concetti. [... essi] sono collegati l'uno con l'altro dai messaggi che si possono scambiare in modo tale che le loro strutture di attivazione sono assai simili agli eventi su grande scala che accadono nel mondo, o che potrebbero accadere in un mondo simile al nostro» (*ivi*, p. 379). Non sfugge la somiglianza di questi pacchetti neuronici con le assemblee cellulari proposte da Hebb circa un trentennio prima. Né si può mancare di notare come la loro relazione di corrispondenza col mondo costituisca una teoria del riferimento grazie alla quale «il significato nasce [...] a causa dell'*isomorfismo*» (*ibidem* [enfasi mia]), isomorfismo che viene definito «infinitamente complesso, sottile, delicato, versatile e intensionale» (*ibidem*). I simboli, inoltre, sono passibili di attivazione e questo causa il passaggio di informazione, cioè la trasmissione di segnali.

A partire da questa caratterizzazione Hofstadter si pone una serie di interrogativi su che cosa effettivamente sia simboleggiato da tali simboli. Essi stanno per elementi o per classi di elementi? In che modo va considerata la loro implementazione neuronale? Disgiunta o sovrapposta? Se sono

sovrapposti come possono essere tra loro distinti? Le risposte che fornisce a queste domande sono molto caute. Ad esempio, egli afferma che la «caratterizzazione dei simboli come "realizzazioni circuitali dei concetti" potrebbe essere nel migliore dei casi una semplificazione eccessiva [... visto che] nello stesso insieme di neuroni possono coesistere parecchi simboli, caratterizzati da configurazioni distinte di attività neuroniche». Inoltre, è necessario sottolineare che «la differenza fra una teoria che contempli simboli fisicamente distinti e una teoria che contempli simboli parzialmente sovrapposti *che si distinguono fra loro per le modalità di attivazione* è che la prima indica una realizzazione dei concetti di tipo hardware e la seconda una realizzazione dei concetti in parte di tipo hardware e in parte di tipo software» (*ivi*, p. 386-387 [enfasi mia]).

Un tipo di realizzazione neuronale sovrapposta conduce a ritenere che i simboli possono essere individuati in maniera univoca all'interno del cervello, ma questo è un aspetto relativamente importante per comprendere i meccanismi del pensiero. Ciò che conta è il modo in cui essi si attivano e, dunque, inviano messaggi. Infatti, Hofstadter afferma che mentre è possibile individuare un simbolo è molto implausibile pensare che esso possa essere preso *isolatamente*, e questo sta a significare che

> l'identità di un simbolo sta proprio nei modi in cui esso è connesso (mediante legami di attivazione potenziale) ad altri simboli. La rete grazie alla quale i simboli sono potenzialmente in grado di attivarsi l'un l'altro costituisce il *modello funzionale* che il cervello si fa dell'universo reale, come pure degli altri universi alternativi che esso prende in considerazione. (*ivi*, p. 390 [enfasi mia])

È interessante notare come nella proposta di un sistema isomorfo col mondo e caratterizzato da una serie di relazioni costitutive della sua valenza semantica si anticipino i temi poi ripresi dalle metodologie connessioniste di una rappresentazione della conoscenza distribuita, anche in modo localistico, cioè con simboli che, in senso puntuale, *corrispondono* direttamente con il mondo. L'introduzione di una relazione di isomorfismo permette di superare gli ostacoli posti dall'iconismo in termini di rappresentazione, giudicato sia non del tutto plausibile per una mente umana, sia a rischio di sintatticismo per un sistema artificiale.

Hofstadter introduce la sua teoria anche allo scopo di approntare una spiegazione di come nella mente umana si possano produrre fenomeni coscienti, argomento che qui non ci è possibile approfondire oltre. Restando su questioni di modellistica simulativa è interessante notare come più oltre in *Gödel, Escher, Bach* egli proponga un'altra definizione di simbolo, più adatta ad essere parte di un programma di IA: «chiamiamo *simbolo* un *frame* che abbia la capacità di generare e di interpretare messaggi complessi» (*ivi*, p. 716). Il simbolo è visto in questo caso come l'unione di una struttura di rappresentazione della conoscenza, il *frame*, e di un attore, cioè una microprocedura capace di produrre un tipo di computazione interattiva all'interno di un sistema computazionale complesso. Se, dunque, come si è visto, il simbolo è anche la realizzazione neuronale di un

concetto, e quindi di fatto corrisponde al concetto stesso, si può concludere che nella teoria che supporta la formulazione dei modelli subcognitivi i concetti non sono mere strutture rappresentazionali, ma in essi rientra costitutivamente e in maniera precipua una parte procedurale che fornisce al concetto la sua funzione attiva. In altri termini, gli elementi semantico-concettuali che possono essere rinvenuti nei modelli subcognitivi, i quali derivano, come si è detto, dalla proposta dell'ipotesi di modello teorico data in GEB, non possono essere rintracciati meramente nell'apparato simbolico dei programmi, bensì è costitutivo della loro natura un aspetto procedurale.

Se nella teoria del sistema mente-cervello proposta in GEB i simboli-concetti hanno una caratterizzazione che li vede fortemente inseriti nella catene di *cluster* neuronali attivi, nei modelli il discorso è ancora di più sbilanciato su un'interpretazione funzionale degli elementi di conoscenza, la quale, di conseguenza, pone in secondo piano la questione della realizzazione fisica su un *hardware* di qualche tipo, perché di fatto irrilevante dal punto di vista del livello esplicativo al quale la teoria dei concetti si pone.

Quale teoria dei concetti, è lecito chiedersi a questo punto, è sottesa all'approccio subcognitivo e alla concezione del mentale che esso intende simulare e spiegare? Questa domanda può essere vista come un imbuto che fa convergere tutti i temi finora discussi. Per avere una risposta occorre procedere per gradi.

È stato fatto notare (Kaplan, Weaver, French, 1990) che sulla base della somiglianza fra assemblee cellulari *à la* Hebb e simboli attivi, questi possono essere considerati alla stregua di «circuiti ricorrenti che forniscono al sistema gli *strumenti* – essenziali, nella nostra visione, ad ogni modello capace di espletare funzioni cognitive – per avere rappresentazioni della realtà interne, semi-autonome, attivabili» (*ivi*, p. 58 [enfasi mia]). Queste strutture sono auspicabilmente individuabili a livello neuronale ma sono funzionalmente strumentali al rappresentazionalismo dei sistemi cognitivi, costituendo in tal modo la condizione *sine qua non* della loro *natura cognitiva*. Tali circuiti ricorrenti (cioè i simboli attivi) sono stati chiamati in vari modi e in molti casi utilizzati per spiegare l'associazionismo concettuale[8]. La tesi suggerita da Kaplan, Weaver e French è che la teoria dei simboli attivi può essere utilizzata per spiegare sia i fenomeni di riconoscimento e categorizzazione, grazie alla corrispondenza instaurata fra circuiti e *cluster* di caratteristiche ambientali fino alla formazione (della rappresentazione) di un concetto; sia i processi cognitivi di alto livello per mezzo della costruzione di reti associative *orientate* di concetti in cui la sequenzialità temporale dei concetti viene rappresentata, appunto, da archi orientati che esprimono l'*ordine* in cui i concetti sono stati associati. A "vicinanze" di entità (eventi o oggetti) nell'ambiente

[8] Il caso più celebre è forse quello di Braitenberg (1984), il quale in merito ai suoi veicoli pensanti afferma che «tutti questi modelli, e altri ancora (basti pensare al fondamentale modello di D. O. Hebb) sono stati creati sulla base dell'idea che il principio fondamentale con cui l'informazione viene elaborata nel cervello è l'associazione, vale a dire il principio secondo cui, quando due cose avvengono insieme, i neuroni che segnalano i due avvenimenti vengono a loro volta collegati da sinapsi. Che queste idea sia corretta, al di là delle prove di natura psicologica, lo hanno provato le ricerche neurofisiologiche degli ultimi decenni» (*ivi*, p. 115). Per una storia degli studi neurofisiologici in merito all'apprendimento e all'immagazzinamento dell'informazione si rimanda a Kandel (2006).

corrispondono “vicinanze” nella rete, che in questo modo si può definire, riprendendo ancora una metafora spaziale, una *mappa cognitiva* dei concetti. Inoltre, se si rende la mappa cognitiva il primo strato di un’ulteriore rete multi-strato è possibile generare processi di astrazione a un meta-livello rispetto a quelli che hanno portato alla formazione della mappa concettuale, fino ad incorporare nella rete come termine ultimo anche i processi di ragionamento logici strettamente dipendenti dalla struttura più che dal contenuto concettuale.

L’idea di un “connessionismo cognitivo” come quello appena esposto è stata a lungo discussa a cavallo fra gli anni ottanta e gli anni novanta, portando alla formulazione di teorie alternative al connessionismo di matrice purista secondo il quale i nodi della rete devono essere equivalenti a neuroni e non rappresentativi di alcunché se presi singolarmente. Ad esempio, Smolensky (1988) parla di livello subconcettuale, indicandolo come più adeguato ad un rappresentazione della conoscenza in un sistema ai fini della spiegazione delle capacità rappresentazionali e delle funzioni cognitive di alto livello del sistema stesso. Il livello subconcettuale si presta senza intoppi alla rappresentaione sia delle caratteristiche (*features*) ambientali su cui si fonda il processo di categorizzazione, sia dei concetti costituenti di concetti più complessi attraverso un’operazione costruttiva di composizione per via associativa, cioè sfruttando il potenziale di strutturazione messo a disposizione dalla rete.

In questo filone vanno inseriti, dal punto di vista della rappresentazione della conoscenza, anche i modelli che abbiamo esaminato nel capitolo precedente. La distinzione fra subconcettuale e subcognitivo pone l’accento sul fatto che il secondo termine si riferisce a modelli che simulano operazioni ad un livello intermedio, mentre il primo a una forma di rappresentazione della conoscenza che, tuttavia, a buon diritto può essere considerata in qualche modo intermedia e tipica dei modelli subcognitivi. Se, infatti, parlare di subcognizione vuol dire riferirsi a meccanismi in grado di attuare tale livello intermedio inconsapevole ma necessario ai livelli cognitivi coscienti – meccanismi che abbiamo visto essere le microprocedure – un altro modo di guardare all’intermediazione fra mente e cervello, cioè tra modelli che simulano l’attività neuronale e modelli che simulano le attività cognitive simbolico-sintattiche, è di farlo attraverso le strutture rappresentazionali conoscitive dei modelli stessi, piuttosto che attraverso le componenti operativo-procedurali, cioè, in definitiva, attraverso i *concetti*: «occorre qualcosa che cada tra questi due livelli di descrizione, molto distanti tra loro. Quello che manca, secondo la mia intuizione [...], è un modello profondo dei *concetti*» (Hofstadter, 1995c, p. 398).

4.6 Modelli dei concetti, concetti come analogie

Lo sviluppo della teoria dei concetti di cui i modelli cognitivi si avvalgono per implementare la TCCL e i processi di ragionamento analogico prende avvio da una rivisitazione del *frame* come

struttura di rappresentazione della conoscenza. In un articolo inedito[9] del 1980, scritto nel periodo di gestazione di SEEK-WHEANCE, Hofstadter, Clossman e Meredith analizzano l'effettivo potere dei *frame* nel dare conto della differenza intensionale/estensionale. Se, infatti, un *frame* può essere visto come un nodo di una rete concettuale, ogni concetto appare avere una descrizione intensionale, cioè una lista di tratti che lo definiscono. Poiché, come è noto, è possibile avere una struttura nidificata di *frame* attraverso il riempimento di *slot* con altri *frame*, diviene centrale la funzione del "puntatore" che traduce in un terminologia algoritmica la nozione di riferimento con l'esterno, cioè con l'estensione del concetto – gli oggetti che ricadono sotto il concetto – e che in una visione che vede i *frame* come nodi di una rete assume il ruolo di collegamento nella rete. I tre autori propongono di considerare una doppia operazione collegata al puntatore (*ivi*, p. 21), quella di "diminuzione di puntamento" (*pointer lowering*), che vede il puntatore diretto verso un oggetto, o verso i "riempitori" (*fillers*) degli *slot* (dunque, un'operazione *verso l'estensione*), e quella di "elevazione di puntamento", attraverso la quale il puntatore è diretto verso un altro *frame*, cioè un nodo astratto ancora passibile di differenti riempimenti (dunque, un'operazione *verso l'intensione*), il quale corrisponde allo scheletro concettuale, cioè all'insieme di concetti-tratti che formano il concetto astratto espresso sotto forma di *ruolo*.

In tale visione è evidente un'esplicita adesione a un rappresentazionalismo simbolico, che viene considerato il giusto livello di descrizione per i fenomeni cognitivi. Il problema del riferimento viene risolto attraverso l'operazione di puntamento verso l'esterno, cioè verso un qualsiasi riempitore. Per tale ragione i nomi di persona vengono considerati «il modo in cui possiamo arrivare più vicini a dare una rappresentazione estensionale di una persona» (*ivi*, p. 22). Da un punto di vista più astratto, attraverso la nozione di "ruolo" viene «generalizzato il concetto formale di *slot*» (*ivi*, p. 26). Infatti, un *frame*-ruolo al posto di uno *slot* sta ad indicare che il riempimento di esso non è univoco, ma implica il «considerare il "significato" del nome dello *slot* in qualche struttura», cioè dal punto di vista del contesto espresso dal ruolo-*frame*, ovvero ancora, dallo scheletro concettuale (l'intensione) che caratterizza quel *frame*.

Al di là degli aspetti più tecnici di questo discorso, si può dire che esso già contenga un tentativo di superamento della tecnica di rappresentazione della conoscenza attraverso *frame* verso forme che ne mantengano gli aspetti positivi, cioè l'idea di un nucleo centrale di tratti condivisi che esprimono l'essenza del concetto e catturano gli effetti di tipicitalità. Le nuove forme di rappresentazione dei concetti sono appunto le reti semantiche peculiari dei modelli che abbiamo visto, le quali permettono di rappresentare in maniera flessibile gli scheletri concettuali, cioè le associazioni di concetti che compongono concetti più complessi. Nelle reti, come si è visto, la gerarchia non è presente in maniera rigida. Esse sono eterarchiche, o, potremmo dire, *dinamicamente gerarchiche*.

[9] L'articolo ha l'ironico titolo: *Shakespeare's plays weren't written by him, but by someone else of the same name* (si veda Hofstadter, Clossman, Meredith, 1980), che ha, però, il pregio di mostrare più di altri come ancora una volta il problema del riferimento, e dunque, del significato, sia stato sentito come centrale nello sviluppo di modelli effettivamente simulativi del pensiero umano.

In aggiunta, ogni nodo concettuale può situarsi a un certo livello di astrattezza ed essere soltanto in maniera mediata riferito ad oggetti esterni al sistema. In tal modo si possono catturare, e rendere esplicite attraverso nodi concettuali appositi, relazioni come quella di somiglianza, che vengono "attivate" (anche se ancora tale termine non viene utilizzato in questa fase di elaborazione della teoria) fra due *frame* che *rispetto a un qualche contesto* sono simili.

In un saggio di poco posteriore (Hofstdater, 1983b), tali idee vengono ulteriormente sviluppate. La questione dei puntatori viene trasformata per mezzo della nozione di slittamento concettuale, slittamento che può essere di due tipi, conformemente all'esposizione iniziale della teoria. Hofstadter individua uno slittamento estensionale, quando una stessa descrizione concettuale si riferisce a due oggetti diversi, e uno slittamento intensionale, quando si passa da una descrizione ad un'altra descrizione del medesimo oggetto (*ivi*, p. 46). Chiaramente il primo si riferisce ai processi di categorizzazione e il secondo a quelli di analogia in senso proprio, nei quali un oggetto assume un diverso ruolo in due contesti. Le nozioni di intensionale ed estensionale vengono sempre viste in modo complementare. Se l'intensione esprime il designatore e si traduce informaticamente nella funzione di puntatore, l'estensione esprime il designato e si traduce informaticamente nell'oggetto, rappresentato da un nodo atomico indivisibile, che entra nel dominio di elaborazione del programma in modo totalmente simbolico (*ivi*, p. 47).

Il problema fondamentale a questo punto è come decidere quale dei due slittamenti è quello più conveniente. La risposta è che dipende dal contesto. È, cioè, il programma a dover decidere nel corso dell'elaborazione. In conclusione, Hofstadter nota che «nel mondo di SEEK-WHEANCE [quello delle successioni dei numeri naturali] la distinzione "intensionale-estensionale" è particolarmente sottile. Forse questo è attribuibile al fatto che qui, le estensioni non sono *oggetti* solidi e tangibili (come pianoforti e corpi umani), ma *concetti* eterei e intangibili (come i numeri e le strutture). Questo significa che nel mondo di SEEK-WHEANCE, le estensioni sono astratte e mentali tanto quanto le intensioni» (*ivi*, p. 50). Nel seguito dello stesso saggio Hofstadter afferma che, oltre al fatto che l'"oggetto" su cui punta l'intensione può essere materiale o meno, le intensioni possono essere in alcuni casi strettamente dipendenti dal contesto, in altri no; possono riferirsi a fatti stabili e permanenti o, al contrario a «connessioni temporanee e forse accidentali» (*ivi*, p. 55).

In definitiva, ciò che egli pone in evidenza è il fatto che non solo si possono dare molteplici, virtualmente infinite, diverse descrizioni intensionali di un medesimo oggetto di qualunque natura esso sia, ma che il modo in cui l'una o l'altra sono rilevanti, sulla base di condizioni differenti, è ciò che veramente determina l'effettivo *significato* dell'oggetto considerato. Questo è, dunque, il vero problema che deve essere affrontato per dotare di "poteri semantici" un programma, o un qualunque sistema artificiale che esibisce funzioni cognitive, e allo stesso tempo è il nocciolo (problematico) in vista di una individuazione dei meccanismi che definiscono il pensiero umano in quanto tale. In altri termini, la problematicità della questione risiede proprio nel modo in cui un sistema intelligente

è un sistema rappresentazionale («un'intensione è – infatti – un elemento di un sistema descrittivo» (*ivi*, p. 47)), il che equivale ancora a dire il modo in cui *riesce a utilizzare in maniera fluida le descrizioni intensionali*:

> Io sono convinto che tutta la flessibilità del pensiero risieda nella fluidità eccezionale dei descrittori intensionali nello slittare in versioni alternative di se stessi sotto la spinta di molteplici pressioni dovute alle circostanze. Io sono dunque convinto che tutte le maggiori intuizioni, sia artistiche che scientifiche, provengono dall'avere il giusto slittamento descrittivo nella giusta direzione a causa del modo in cui le pressioni esterne si sono accumulate. (*ivi*, p. 55)

Di qui, il passo all'implementazione di reti concettuali è breve. L'utilizzo di connessioni dinamiche permette lo slittamento e la flessibilità. Il richiamo al connessionismo riguarda, dunque, questo punto, cioè il modo in cui la conoscenza viene rappresentata, anche se attraverso un simbolismo locale e non un subsimbolismo distribuito di fatto a-simbolico.

La discussione del rapporto fra estensione e intensione dà adito ad alcune osservazioni. In primo luogo, essa si conclude con una affermazione netta del rappresentazionalismo nella spiegazione dei fenomeni cognitivi, il quale si mantiene come aspetto esplicativo condiviso da tutti i modelli subcognitivi. Secondariamente, essa è un preludio alla definizione del problema del *grounding* dei simboli trattati da un sistema artificiale i quali rappresentano la sua conoscenza (Harnad, 1990, 2003). Laddove Harnad indica come ineliminabile un "ancoraggio a terra" dei simboli del sistema[10], Hofstadter pone, con un decennio di anticipo, la questione del riferimento, del rimando fra simboli attraverso gerarchie ricomponibili, facendone un punto cruciale dello sviluppo di sistemi intelligenti. Infine, il richiamo alla distinzione estensionale e intensionale indica ancora una sorta di ambiguità nel trattare il tema della rappresentazione della conoscenza e del rappresentazionalismo della mente. Infatti, Hofstadter parla di descrizioni intensionali, rappresentate dal nodo-*frame*, ma ancora espresse in linguaggio naturale, tralasciando di porre una distinzione netta fra *concetti* e *parole* anche se i "nomi", intesi come termini singoli e come locuzioni e proposizioni, vengono considerati perlopiù *etichette* che permettono il riferimento. Tuttavia, resta l'ambiguità dovuta alla sovrapposizione fra concetti, come strutture rappresentazionali e *strumenti* di conoscenza, ed elementi del linguaggio naturale (termini e proposizioni), ambiguità che esce rinsaldata dalla

[10] Non c'è spazio in questa sede per discutere le idee di Harnad. La sua posizione potrebbe essere riassunta dalle seguenti parole: «in un sistema simbolico intrinsecamente dedicato ci sono più vincoli sui segni (*token*) dei simboli che quelli meramente sintattici. I simboli sono manipolati non solo sulla base della forma (*shape*) arbitraria del loro segno, ma anche sulla base della "forma" decisamente non arbitraria delle rappresentazioni iconiche e categoriali connesse ai simboli elementari tenuti a terra (*grounded*) da cui sono composti i simboli di ordine più alto. Di questi due tipi di vincoli, quelli iconici/categoriali sono preminenti» (Harnad, 1990, p. 342). La soluzione del problema del *symbol grounding* è vista da Harnad, dunque, nell'utilizzo di reti connessioniste correlate direttamente e *non arbitrariamente* allo stimolo che categorizzano. Tuttavia, così come l'iconismo, anche la relazione di covarianza non è scevra da problemi. Per una discussione ancora valida di tali questioni dal punto di vista filosofico si rimanda a Cummins (1989).

trattazione del problema in termini di contrapposizione intensionale-estensionale, ancorché nella prospettiva di un suo superamento.

Con lo sviluppo dei modelli e il consolidarsi della teoria dei concetti tale ambiguità rimane. Se l'utilizzo di descrizioni linguistiche allo scopo di mostrare la loro variabilità di significato in dipendenza dal contesto è stata all'inizio funzionale alla spiegazione dei meccanismi creativi di spostamento fra descrizioni alternative di un medesimo oggetto (di conoscenza, senza riguardo alla sua materialità), il quale è in grado di assumere differenti ruoli[11], il distacco della teoria dei concetti dagli aspetti linguistici si accentua in seguito. Ad esempio, intervenendo nel dibattito sulle idee di Smolensky e la controversia "tra simboli e neuroni" (Smolensky, 1988), Hofstadter rivede la nozione di "sfera controfattuale implicita" (Hofstadter, 1985d, 1985e) in quella di "alone concettuale".

Secondo questa accezione ogni concetto è circondato da un *alone concettuale* che esprime i concetti correlati ad esso, o, altresì, se si considera una determinata situazione, le sue alternative, i mondi possibili, i controfattuali ordinati secondo un grado crescente di lontananza dalla situazione iniziale. È abbastanza intuitivo come l'alone delle situazioni serva a rappresentare la conoscenza di senso comune. Dal punto di vista dei concetti la nozione di "alone" costituisce l'aspetto centrale della teoria dei concetti hofstadteriana. Esso, infatti, «è distribuito e non ha confini precisi. [È un] prodotto inevitabile e epifenomenico della "topologia mentale" – cioè una visione dei *concetti in quanto intrinsecamente distribuiti*, visti come *regioni che si intersecano in uno spazio*» (Hofstadter, 1988, p. 159-160). Lo stesso Hofstadter riconosce che «c'è poco di originale in tutto ciò – è solo un modo di dire che la mente è strutturata in modo associativo» (*ibidem*). Tuttavia, pagato anche il tributo all'associazionismo, la teoria risulta interessante per il modo in cui sviluppa la sua versione di associazionismo, soprattutto dal punto di vista simulativo. Infatti, aspetto fondamentale, «quando i *concetti* sono adeguatamente rappresentati in un modello (cioè sono rappresentati come regioni sovrapposte in uno spazio astratto) gli aloni concettuali sono automaticamente presenti; non c'è bisogno di aggiungere al modello alcun apparato» (*ibidem*). L'"adeguata rappresentazione" di cui Hofstadter parla è la maniera ricca e particolareggiata in cui i modelli subcognitivi hanno implementato questa teoria dei concetti, in stretto legame con quella del *loop* centrale cognitivo. La sua adeguatezza costituisce proprio la parte problematica del problema, e perciò la più interessante. Se ancora nelle poche pagine del 1988 Hofstadter scriveva che il suo scopo era stato quello di «far risaltare la stretta relazione che intercorre fra senso comune ed un'architettura connessionista (o almeno associazionista) del mentale» (*ibidem*), con una propensione manifesta a favore delle architetture «subsimboliche», negli anni seguenti abbiamo visto che l'adesione al connessionismo ha un certo numero di distinguo e consiste in definitiva, anche stando ai modelli sviluppati,

[11] Si veda anche Hostadter (1985d), la cui versione originale data 1982 e in cui è ancora presente una "presentazione linguistica" dei concetti, unitamente ad affermazioni sul ruolo centrale della variazione per i processi di pensiero: «*lo slittamento non intenzionale ma non accidentale permea i nostri processi mentali, ed è il vero punto cruciale del pensiero fluido*» (*ivi*, p. 237).

nell'acquisizione dell'idea *e* di una conoscenza *e* di un'elaborazione distribuite, ma ancora nel campo del simbolico. Che cosa possiamo concludere, dunque, in merito alla teoria dei concetti implementata attraverso i modelli subcognitivi?

La centralità del ruolo dei concetti per lo studio della cognizione viene decisamente rivendicato all'interno dell'approccio subcognitivo: «il germe della cognizione umana sono i concetti. Io credo che prima di poter fare progressi fondamentali nella comprensione della cognizione umana, dobbiamo capire molto di più in merito ai concetti: come *si sviluppano*, come *evolvono* e come *influenzano altri concetti*» (French, 1995, p. 180 [enfasi mia]). Se nelle microprocedure va visto il modo in cui i livelli intermedi di elaborazione (e, conseguentemente, del pensiero) posti fra quello simbolico e quello neuronale vengono implementati, indipendentemente dal problema lasciato aperto della loro giustificazione cerebrale, i concetti sono la controparte di questa "via di mezzo" dal punto di vista della conoscenza. Abbiamo visto come nei modelli subcognitivi essi siano rappresentati in uno specifico modulo dell'architettura triadica che mette in atto il *loop* centrale cognitivo. Ogni concetto non è visto come un'unità compatta, ma in maniera attiva, di modo che non siano colte solo le «proprietà *statiche* dei concetti – per esempio, i giudizi indipendenti dal contesto di appartenenza a una categoria – bensì anche il modo in cui i concetti si allungano, si piegano e si adattano alle situazioni impreviste» (Hofstadter, 1995b, p. 331).

I concetti modellati, perciò, corrispondono pienamente ai simboli attivi, nel senso che la loro rappresentazione non si riduce alla semplice presenza di un nodo nella rete. Così come i simboli attivi erano l'unione di una struttura di conoscenza e degli agenti ad essa collegati, le microprocedure sono indispensabili ad un modellamento dei concetti come aspetti principali del mentale. Lo scollamento fra concetti della mente e concetti nei modelli è solo apparente, se si considerano i primi come *risultato emergente* dell'elaborazione dei modelli subcognitivi a partire da alcuni vincoli prefissati. Questi vincoli sono quelli espressi dalla rete. Si è visto come molti modelli, soprattutto fra quelli meno recenti, non prevedano l'introduzione di nuovi concetti come nodi. Inoltre le connessioni sono sì variabili quanto alla lunghezza, ad indicare il mutamento di forza del legame associativo prodotto dall'apprendimento, ma nuove connessioni in genere non vengono predisposte. Tuttavia, questi sono problemi che una volta sarebbero stati definiti "empirici" e che oggi si possono chiamare d'implementazione. Infatti dal punto di vista del modello la nascita di nuovi concetti non è preclusa, grazie proprio all'elaborazione emergente. L'attivazione di *cluster* di concetti nella rete è il corrispettivo della nascita di concetti complessi a partire da quelli più semplici per composizione e senza che ci siano vincoli gerarchici imprescindibili, soprattutto per quanto riguarda i concetti più astratti (si pensi all'esempio di "la successione delle identità" e "l'identità delle successioni"). Se, infatti, un concetto non deve essere per forza espresso da un termine, ma è concetto anche ciò che è etichettato da un'intera espressione linguistica, i concetti prodotti dall'elaborazione nei modelli simulativi sono molti e in alcuni casi passibili di memorizzazione attraverso la creazione di apposite strutture rappresentazionali, come ad esempio si

è visto in METACAT. Inoltre, se è l'elaborazione a produrre l'emergenza dei concetti e una descrizione in termini linguistici non è necessaria, il modo ultimo di individuazione dei concetti è proprio quello di considerare le *azioni* che essi provocano nel corso dell'elaborazione. Una visione procedurale, più che oggettuale, della conoscenza conduce, dunque, a un criterio pragmatista (*á la* Peirce) di individuazione dei concetti.

Se non si vuole scomodare la teoria pragmatista del significato con i suoi "effetti concepibili", la quale pure è quasi imprescindibile dal punto di vista simulativo – indipendentemente da come la si intenda implementata in un modello – perché forte del debito contratto con la necessaria valutazione della prestazione del modello, basterà dire che anche i concetti possono essere considerati dal punto di vista *funzionale* proprio per il fatto di essere attivi. Tuttavia, in relazione alla rappresentazione, l'aspetto cruciale risiede nella loro natura «semi-distribuita, poiché un concetto nella Rete di Slittamento è distribuito probabilmente soltanto su un piccolo numero di nodi: un nodo centrale e il suo alone probabilistico di slittamenti potenziali» (Mitchell, 1993, p. 226).

Il fatto che sia possibile, *contestualmente*, considerare ogni concetto come primitivo, cioè come termine ultimo della struttura gerarchica specifica di conoscenza che entra nel campo di un'elaborazione particolare, aggiunge un dettaglio molto importante ai fini dell'esatta comprensione della natura *semi-distribuita* dei concetti. Infatti, l'essere semi-distribuiti è interpretabile in una doppia direzione. In altre parole, un concetto non va visto solo come una parte fissa e una parte mobile o variabile costituita dall'alone soggetto ai mutamenti di attivazione nella rete, allo stesso modo, per utilizzare una metafora, di un perno cui sono collegati molteplici ingranaggi semoventi e in continua trasformazione. Piuttosto, la parte *non* distribuita del concetto nella teoria può essere *alternativamente* vista nel nucleo del concetto, e dunque rappresentata nella memoria dei modelli dal nodo corrispondente, *o* nell'alone stesso, una volta fissato il quale è il nucleo centrale a godere della possibilità di muoversi, di mutare. In questo va vista l'essenza dello *slittamento*, che è produttivo di nuova conoscenza. In conclusione, si può affermare che è proprio la natura semi-distribuita dei concetti a garantire ai modelli un'elaborazione simbolica dell'informazione anche se non esclusivamente basata su manipolazioni sintattiche. Questo perché, se la composizionalità dei concetti è permessa dalle proprietà costruttive dell'attivazione congiunta di cui gode una rappresentazione in forma di rete, l'elaborazione concettuale, che comporta il passaggio da un concetto all'altro e che costituisce il nucleo del pensiero autonomo e creativo, è sostanziata dalla rappresentazione dinamicamente mutevole dei concetti, fatto salvo il fondamentale vincolo per cui *almeno uno dei due fra il nucleo del concetto e il suo alone concettuale, strettamente connesso al contesto globale della rete, devono essere fissati.*

La rappresentazione semi-distribuita dei concetti nei modelli subcognitivi intende, inoltre, catturare alcune degli aspetti dei *concetti* considerati essenziali sia dalle ricerche in psicologia che dalla riflessione filosofica. Consideriamoli separatamente.

Dal punto di vista psicologico, appartiene ormai alla storia della psicologia la profonda revisione, non priva di contrasti, cui è stata sottoposta quella che viene definita la "teoria classica dei concetti". In termini generali, si può affermare che, secondo questa teoria, comunemente fatta risalire a una matrice di stampo filosofico (a partire da Platone e Aristotele)[12], i concetti sono considerati definizioni che individuano tutto e soltanto l'insieme degli oggetti che ricadono all'interno dell'estensione del concetto. Se un oggetto possiede la lista di tratti che costituiscono l'intensione del concetto è compreso nel suo campo di applicazione; in caso contrario no. Non c'è discrezionalità. L'attribuzione categoriale è netta. Un cambiamento di questa prospettiva si è reso necessario a seguito dell'ampio numero delle evidenze sperimentali raccolte dalla psicologia negli ultimi cinquanta anni, le quali hanno mostrato come gli esseri umani non si comportino in maniera netta nell'applicazione delle categorie. Molte teorie alternative sono state proposte a cominciare da quella dei "concetti come prototipi" o, più correttamente, "teoria prototipica dei concetti", avanzata dalla Rosch[13]. In base a questa teoria, ogni concetto è costituito da un insieme di caratteristiche (*feature*) pesate. A ogni caratteristica corrisponde in valore che è tanto più grande quanto più si ritrova nei membri riconosciuti della categoria. In sostanza, l'attribuzione categoriale equivale al computo della somma dei pesi delle caratteristiche di uno stimolo e il superamento di un determinato valore di soglia ne causa l'attribuzione. Come è facile intuire, tale teoria intende dare conto degli aspetti di *tipicalità* riscontrabili nei concetti, spiegando allo stesso tempo la maggiore velocità di attribuzione per i membri tipici di una categoria. L'idea di un computo di valori a superamento di soglia permette di identificare i concetti anche in base alla loro vicinanza o lontananza concettuale e di cogliere gli aspetti sfumati della delimitazione categoriale. Tale teoria, peraltro, non è esente da problemi. Ad esempio, essa si adatta meglio a concetti che riguardano categorie naturali, mentre mostra i suoi limiti con i concetti di manufatti o con quelli più astratti.

Per risolvere questo problema è stata proposta una teoria detta del "*core* più prototipo" o del "nucleo più procedure di identificazione" (Miller, Johnson-Laird, 1976) o, più direttamente, "binaria" (Hampton, 1988). Il duplice risvolto della teoria risiede nel fatto che un concetto è identificato con un nucleo essenziale, cui si aggiungono una serie di procedure per individuare le caratteristiche superficiali che ne indicano il grado di tipicità e determinano la gradualità dell'appartenenza categoriale. Anche questa teoria è stata criticata, soprattutto per quanto riguarda il suo essenzialismo, inteso come l'aspetto sfuggente legato all'impossibilità di determinare in via sperimentale l'esistenza del nucleo che costituisce uno dei due corni in cui viene scisso il concetto.

[12] Un adattamento della teoria dal punto di vista psicologico è in Bruner, Goodnow, Austin (1956).

[13] I riferimenti sono molteplici. Per un'esposizione generale della teoria si rimanda a Rosch (1975). È stato fatto notare in sede di ricostruzione storica (Murphy, 2002) che non del tutto sorprendentemente la teoria della Rosch fu all'inizio interpretata in modo erroneo, il prototipo di una categoria venendo considerato come l'esempio migliore di tutti gli appartenenti alla categoria, e non invece come l'insieme della caratteristiche tipiche della categoria. La differenza è sottile ma rilevante, ed è possibile che l'errata interpretazione sia stata anche frutto di un accostamento eccessivo con una versione semplificata della teoria, filosofica, delle somiglianze di famiglia formulata da Wittgenstein, la quale, invece, sembra anche più conforme alla teoria dei concetti come analogie che vedremo in seguito.

Un altro modo di risolvere i problemi posti dalla teoria dei prototipi è stato quello di pensare a una possibile strutturazione della lista di tratti prototipici attraverso l'introduzione di *schemi* organizzativi di questi tratti (Rumelhart, Ortony, 1977). In tal modo le caratteristiche diventano tipi che possono essere istanziati da un insieme ristretto di differenti valori. Inoltre, questi tipi, pensabili come *slot* di un *frame*, possono esprimere anche relazioni, in modo da risolvere il problema lasciato aperto dall'affermare semplicemente che un concetto consiste in una lista di tratti pesati, cioè più o meno prototipici, senza ulteriori specifiche. In questa proposta appare evidente come teoria dei prototipi e strutture di rappresentazione della conoscenza quali sono i *frame* sono strettamente imparentate, anche se il passaggio dalla prima alle seconde comporta alcune revisioni della teoria attraverso l'immissione di aspetti legati alle costanti relazionali strutturali e alle relazioni concettuali vincolate come quelle fra tipi e istanze. Ciò conduce, tuttavia, a un lungo discorso sulle gerarchie concettuali che non è possibile sviluppare appieno. Sta di fatto che, a livello generale, è possibile considerare la rappresentazione strutturata dei concetti come un primo passo verso il superamento della teoria dei prototipi, superamento, è questo il punto interessante, innescato da riflessioni esterne al campo ristretto della psicologia, poiché influenzate dagli sviluppi dell'IA.

Un'altra teoria alternativa a quella dei "concetti come prototipi" è quella dei "concetti come collezioni di esempi" (Nosofsky, 1988) che deriva dalla "teoria del contesto di classificazione" (Medin, Schaffer, 1978). Questa teoria, di stampo più olistico delle precedenti, supera definitivamente l'idea di una lista di tratti, che abbiamo visto essere deterministica nel caso della teoria classica e statistica nel caso della teoria dei prototipi. I concetti vengono considerati collezioni di esempi ed ogni nuovo input percettivo viene classificato attraverso un certo concetto in base alla stima della somiglianza o meno con gli esempi di cui il concetto è costituito. Tuttavia, anche in questo caso si pongono dei problemi, il primo dei quali riguarda il modo in cui gli esempi vengono immagazzinati in memoria e in cui fronteggiare questo ingente carico di informazione.

Altre teorie di diverso tipo sono state proposte, teorie che valutano la componente relazionale e funzionale dei concetti più che la loro struttura. Tra queste la più nota è la "teoria dei concetti come teorie" (Murphy, Medin, 1985; Gopnik, Meltzoff, 1997), che sottolinea il fondamentale contributo della conoscenza generale del dominio nel processo di attribuzione categoriale. In altri termini, è la nostra conoscenza che abbiamo sul mondo, le nostre teorie, a determinare il modo in cui un concetto viene applicato. E le teorie sono più che liste di tratti, poiché comprendono anche un sistema di relazioni fra le loro parti, relazioni che entrano inevitabilmente nel processo di categorizzazione. Un'altra teoria "anti-strutturale" dei concetti è la teoria dei concetti *ad hoc* o *goal-oriented*", sostenuta da Barsalou (1983), secondo il quale molti concetti, se non tutti, vanno considerati in modo funzionale, risiedendo la loro natura nel fine per cui vengono creati e dipendendo in tal modo dalle esperienze e dagli scopi di chi li utilizza per la categorizzazione. Un

teoria collegata a questa è quella, già accennata nel precedente capitolo, dei concetti dipendenti dall'azione[14].

Questo breve[15] *excursus* fra le teorie psicologiche dei concetti dovrebbe mmeterci nella condizione di determinare la portata della teoria dei concetti implementata nei modelli subcognitivi. Abbiamo visto come in tali modelli sia presente una progressiva evoluzione, dominio-dipendente, verso la simulazione di processi categorizzazione basati su primitive percettive che si aggiungono alle primitive relazionali tipiche della simulazione subcognitiva fin dai suoi inizi. In Foundalis (2006) il processo è portato alle estreme conseguenze, tanto che la simulazione dei concetti richiama esplicitamente la teoria dei prototipi e quella degli esemplari delle quali viene suggerita una fusione attraverso l'impiego del *General Context Model* (Nosofky, Palmeri, 1997), un insieme di formule per calcolare il grado di similarità fra due esempi allo scopo di attuare la *pattern formation*. Tuttavia, a livello generale va osservato che i processi di categorizzazione sono soltanto uno degli aspetti che entrano nella simulazione messa in atto dai modelli subcognitivi, un aspetto che, peraltro, si aggiunge a quelli da cui tutta l'impostazione prende le mosse.

La modellizzazione dei concetti attraverso le reti semantiche di slittamento intende cogliere gli aspetti di tipicalità, grazie ad opportune topologie delle reti. Inoltre, la grande importanza riservata al contesto fa sì che un nodo concettuale e il suo alone non siano una semplice implementazione della teoria binaria, ma, piuttosto, *ad un certo livello*, modellizzazioni delle funzioni spiegate da teorie come quella dei concetti come teorie o dei concetti *ad hoc* e *goal-oriented*. Per quanto riguarda gli effetti legati alla tipicalità, si può dire che le reti concettuali, unitamente ai meccanismi elaborativi, modellino efficacemente i confini sfrangiati fra concetto e concetto, e allo stesso tempo un nucleo centrale che è differente dal *core* della teoria binaria, perché individuato dal contesto ristretto delle attivazioni concettuali e da quello più largo dell'elaborazione microprocedurale del sistema da cui esse dipendono.

La ricostruzione degli sviluppi della rappresentazione della conoscenza nei modelli subcognitivi mostra un progressivo raffinamento del modo in cui la conoscenza permanete viene implementata nelle reti semantiche, da ultimo anche attraverso forme di *learning* supervisionato. Il ruolo rivestito dalle teorie psicologiche dei concetti in questo processo è manifesto. L'influenza che esse hanno avuto ha contribuito a rendere tali modelli qualcosa di più di semplici simulazioni di meccanismi associativi. Inoltre, seppure il *background* rimanga quello filosofico dell'associazionismo di matrice empirista e pragmatista, la modellizzazione della conoscenza risente degli sviluppi paralleli compiuti dal connessionismo nel campo della rappresentazione della conoscenza. Le reti di concetti semi-distribuiti contribuiscono all'affermazione di quest'ultimo, condividendo molteplici aspetti con le rappresentazioni distribuite del connessionismo, in particolare con le reti neurali localistiche in cui ad ogni nodo corrisponde un concetto. La distanza con queste rimane, però, nel fatto che la

[14] Si veda Borghi (2002).
[15] Si rimanda a Borghi (1996) e a Murphy (2002) per una rassegna e una discussione dettagliate delle teorie sui concetti dal punto di vista psicologico (e non solo).

conoscenza dei modelli cognitivi non risiede esclusivamente nella rete, ma nell'interazione globale dinamica della parti costitutive dell'architettura. Non ci sono algoritmi che regolano *in proprio* l'attività della rete semantica secondo le regole del connessionismo. La conoscenza delle reti concettuali nei modelli subcognitivi del FARG è tale *esclusivamente in virtù delle microprocedure*, che contribuiscono a renderla attiva. Resta, tuttavia, irrisolto il problema del riferimento di questa conoscenza a meno di non postulare, come si è visto nei modelli più legati ad aspetti percettivi di basso livello, l'innata capacità di recepire caratteristiche (*feature*) "ambientali" non arbitrarie, cioè *esattamente corrispondenti* a micro-dispositivi mentali approntati per afferrarle. In una concezione strutturale dei concetti questo limite sembra essere l'unica garanzia per un riferimento *stabile* alla realtà, così come in parte era stato già suggerito da Harnad (1990) in merito alle condizioni ipotizzate per il *symbol grounding*. Tale prospettiva di saldatura col livello percettivo più basso costituisce, dunque, una delle principali questioni cui la futura ricerca deve tentare di dare una risposta.

Finora ci siamo occupati delle teorie psicologiche dei concetti. Che cosa possiamo dire, in via preliminare all'esposizione della teoria dei concetti che supporta i modelli subcognitivi, delle relazioni che tali modelli intrattengono con le teorie filosofiche dei concetti?

Non è possibile affrontare in questa sede neppure una succinta esposizione delle teorie filosofiche in merito ai concetti considerata l'ampiezza del tema, il quale da un certo punto di vista si distende lungo tutta la storia del pensiero occidentale. Anche limitandoci alle teorie più recenti, frutto della riflessione filosofica novecentesca e in parte debitrici delle ricerche sviluppate nel campo della psicologia e dell'IA, il compito è fin troppo grande[16]. In linea generale, si può dire che il ruolo rivestito dal linguaggio nelle teorie filosofiche dei concetti è preponderante, al punto che spesso concetti e termini del linguaggio vengono sovrapposti e utilizzati in sede di argomentazione in maniera interscambiabile. Inoltre, solo negli ultimi decenni si è arrivato a distinguere, anche se non sempre, fra pensiero e linguaggio e a considerare i concetti come costituenti del pensiero che possono spiegare l'agire intenzionale degli esseri umani intesi come sistemi intelligenti. In questa sede, prenderemo in considerazione quattro proprietà dei concetti, la cui spiegazione è ritenuta essere una condizione *sine qua non* di ogni teoria dei concetti dal punto di vista filosofico e valuteremo la loro effettiva implementazione nei modelli esaminati nel corso di questo lavoro per poi passare alla discussione della teoria che li supporta.

Coliva (2004) ritiene che ogni teoria dei concetti deve poter spiegare la loro 1) composizionalità; 2) pubblicità; 3) efficacia causale; 4) normatività.

Per quanto riguarda la *composizionalità*, essa sembra soddisfatta dal carattere strutturato dell'implementazione della conoscenza nei modelli. Tale proprietà, infatti, indica il potere produttivo dei concetti, soddisfatto nei modelli dall'associazione dei concetti tramite attivazione nel

[16] Per una rassegna della principali teorie filosofiche dei concetti si veda Coliva (2004).

corso dell'elaborazione, associazione che si attua per gradi ed è virtualmente illimitata (limitata solo dalle risorse computazionali). Inoltre, i concetti nei modelli sono implementati tenendo conto anche della loro sistematicità, attraverso l'utilizzo di meta-nodi concettuali che esprimono relazioni di simmetria tra concetti. Tali relazioni, come quella di opposizione o successore, permettono il dispiegarsi di una rappresentazione concettuale tendente alla coerenza sistematica al fine di giungere ad una visione unitaria e ad una via di uscita univoca dall'elaborazione. Per tale ragione, i meta-nodi relazionali vanno visti come necessari all'elaborazione stessa, in quanto permettono di organizzare la situazione percepita secondo schemi precisi, i quali in qualche modo ricalcano le relazioni spaziali. Essi devono essere considerati uno degli assiomi della teoria, pena l'impossibilità di attuare schemi concettuali coerenti. La loro esistenza come puntello di ogni meccanismo di ragionamento è un postulato.

In merito al requisito della *pubblicità*, esso è garantito sia dal simbolismo insito nella forma scelta di rappresentazione della conoscenza, sia dall'elaborazione basata sul rinvenimento di primitive relazionali nello stato di cose analizzato. Tuttavia, il parlare di pubblicità relativamente a sistemi artificiali che modellano la conoscenza concettuale è questione strettamente legata a quella del riferimento dei simboli, che almeno in parte deve trovare una radice comune nel fatto che sono i programmatori umani a decidere i concetti della rete semantica. I concetti nel programma e quelli umani sono *a fortiori* co-referenziati se si parla di quelli immessi come base permanente dal programmatore stesso; sono condivisi fra sistema e utente se si accetta l'architettura del programma come valida spiegazione dei processi mentali. Il problema, dunque, è di ordine metodologico ed epistemologico, e riguarda i principi della particolare forma di simulazione adottata.

È evidente come nella prospettiva simulativa impiegata in questo caso specifico i concetti abbiano un'*efficacia causale*. Essi, infatti, non vanno visti soltanto come una rappresentazione statica di conoscenza, ma inseriti con un ruolo di guida nell'elaborazione attraverso le pressioni *top down* messe in atto sia singolarmente sia globalmente dalle microprocedure. In questo sta l'aspetto eminentemente funzionalista della loro implementazione, il quale è visibile nei processi di alto livello di creazione di strutture rappresentative e di analogie messi in atto dal programma attraverso l'interazione fra le parti dell'architettura. Da questo punto di vista si può dire che il loro ruolo funzionale sta *intrinsecamente* nelle potenzialità *attive* della rete ed *estrinsecamente* nelle attività delle microprocedure, da cui dipende l'attività della rete.

Infine, le *proprietà normative* dei concetti ancora una volta sono implementate attraverso la particolare topologia che viene data alla rete e che stabilisce le associazioni concettuali, per cui i concetti possono essere posti in un legame gerarchico che esprime l'inclusione categoriale e le relazioni di mutua esclusione fra concetti di oggetti appartenenti alla stessa categoria (come si è visto soprattutto nella complessa rete semantica di TABLETOP). Inoltre, la particolare natura interattiva della rete con gli elementi percepiti, dovuta allo scambio informazionale fra le due componenti dell'architettura permette di stabilire in maniera quantitativa l'esatta influenza di ogni

concetto, espressa dalla sua attivazione, sul comportamento del programma. Ancora una volta questa è una conseguenza di uno schema implementativo generale mirante in primo luogo a simulare la natura composizionale e causale dei concetti e solo secondariamente i processi di categorizzazione a partire da stimoli percettivi di basso livello.

Quale teoria dei concetti può essere vista supportare questo tipo di implementazione? La proposta fatta da Hofstdater, in stretto legame con la TCCL e con il modello generale di architettura ad essa legato, è quella di considerare «un concetto come un pacchetto di analogie» (Hofstadter, 2001, p. 507). In atri termini, «ogni concetto che possediamo è niente altro che un *fascio* ben impacchettato di analogie; tutto ciò che facciamo quando pensiamo è muoverci in maniera fluida da concetto a concetto – il che equivale a dire, saltare da un *fascio* di analogie ad un altro – e, inoltre, tali salti da concetto a concetto sono essi stessi compiuti attraverso connessioni analogiche» (*ivi*, p. 500 [enfasi mia]). In tal modo, Hofstadter arriva a saldare i vari aspetti dell'analogia. Infatti, secondo questa teoria, che potremmo denominare "teoria dei concetti come analogie", i concetti esprimono punti di convergenza degli stimoli esterni *dello stesso tipo* in conformità al processo di categorizzazione. Allo stesso tempo, il passaggio da un concetto all'altro avviene per via analogica, nel senso che, come si è visto nei modelli, oltre a essere influenzato dagli elementi della situazione percepita, avviene all'interno di pressioni contestuali (della rete) che portano a considerare il ruolo rivestito da un concetto all'interno di uno schema (scheletro) concettuale come passibile di occupazione da parte di un altro concetto.

La saldatura fra processi di categorizzazione e di costruzione di analogie in senso proprio sotto una nozione più ampia del termine "analogia" passa attraverso la teoria dei concetti come analogie, che in definitiva può riassumersi nell'idea che esiste sempre un contesto che *recupera* gli elementi ad esso più adattabili, siano esse rappresentazioni mentali (simboliche) direttamente collegate agli stimoli percettivi in un legame di non arbitrarietà come quello tra *feature* e corrispondente rappresentazione atomica, o rappresentazioni più complesse costituite da sistemi concettuali organizzati secondo uno schema di relazioni, la cui autonomia rende il sistema che detiene una tale capacità operativo in modalità *off-line*. Secondo Meini e Paternoster, sono due le principali caratteristiche di questo tipo di modalità: «l'abilità di attivare rappresentazioni in una maniera *top down*, senza richiedere la presenza di uno stimolo (questa è la capacità di *distaccamento dal lato dell'input*); l'abilità di non attivare, in presenza di uno stimolo dato, l'azione (complessa) che di solito è attivata dallo stimolo (questo è la capacità di *distaccamento dal lato dell'output*, ovvero la capacità di *inibire* un'azione)» (Meini, Paternoster, in corso di pubblicazione). Entrambe queste abilità sono implementate nei modelli subcognitivi. La prima riguarda la natura funzionale dei concetti; la seconda la capacità di modificazione dinamica della rete alla base dei processi autonomi del sistema, attraverso le influenze di ritorno sull'insieme delle microprocedure operative. Esse, inoltre, sono anche spiegate dalla teoria dei concetti come (pacchetti di) analogie proprio per il fatto che essa mira a descrivere il passaggio fluido, autonomo e creativo da un concetto all'altro.

Se ormai dovrebbe essere chiaro il ruolo ineliminabile del contesto nel modo in cui l'attivazione concettuale e la concettualizzazione funzionano a tutti i livelli, appare chiaro che il contesto di per sé non può spiegare tutto. Occorrono, per così dire, dei puntelli, che, come abbiamo visto discutendo delle reti concettuali semi-distribuite che implementano i concetti come nuclei circondati da aloni, impediscano all'elaborazione, cioè ai processi mentali di ragionamento, di perdersi in una continua sequenza di rimandi privi di utilità cognitiva. Infatti, se il passaggio analogico di concetto in concetto spiega le modalità di elaborazione *off-line*, soprattutto per quanto riguarda la formulazione di controfattuali, le "variazioni sul tema", occorre che il processo si appoggi da qualche parte. Hofstadter ci dice che il processo di ragionamento analogico in senso lato può essere anche considerato da un altro punto di vista, quello dei suoi, potremmo chiamarli, "dispositivi di stabilità", ovvero gli

> "attrattori percettivi", *loci* della memoria a lungo termine che vengono ingranditi quando si incontrano le situazioni [che li richiamano]. Noi tutti abbiamo molte migliaia di questi attrattori nelle nostre memorie dormienti, a una minuscola frazione delle quali soltanto abbiamo accesso quando incontriamo una nuova situazione. (Hofstadter, 2001, p. 522)

E in maniera molto interessante aggiunge:

> Da dove scaturiscono questi attrattori? Quanto sono pubblici? Possiedono espliciti indicatori o etichette? Eccone una lista dei tre tipi principali:
>
> - item lessicali standard (parole, nomi, frasi, proverbi, ecc.) forniti a un ampio pubblico attraverso un ambiente linguistico condiviso;
> - esperienze altrui, diffuse in un pubblico vasto attraverso i mezzi di comunicazione (cioè, luoghi, personaggi ed eventi di piccola e grande scala in libri, film, show televisivi, e così via), la più piccola delle quali ha un'etichetta linguistica esplicita e la più complessa delle quali non ne ha nessuna;
> - memoria personali uniche, mancanti di ogni etichetta linguistica prefissata (tali pezzi sono generalmente molto grandi e complessi, come ricordi di un lontano passato, o perfino eventi che si dispiegano in un tempo assai lungo, come il corso preferito alle superiori, un anno speso in una città speciale, un divorzio protratto e così via). (*ivi*, pp. 522-523)

L'idea che la stabilità del pensiero sia garantita da attrattori o punti di attrazione nella memoria è esplicitamente ricondotta da Hofstadter a Kanerva (1988) e al suo lavoro sulle memorie distribuite. Tutto ciò collima con la rappresentazione della conoscenza in una rete in cui i concetti sono semi-distribuiti: il nucleo centrale del concetto costituisce l'attrattore. Tuttavia, non ha senso pensare al nucleo centrale privato del suo alone modificabile. Senza entrambe le componenti si perderebbe

tutto il potere analogico, che risiede nella possibilità di slittare da un (nucleo centrale di un) concetto ad un altro, ovvero da un attrattore ad uno "vicino" attraverso il contesto. Tale contesto espresso dall'alone va dunque pensato come locale ma anche interrelato con l'intera rete. Infatti, ogni nodo ha più collegamenti con gli altri nodi e dal punto di vista strutturale i pezzi della rete, quelli che Leibniz vedrebbe se essa forse all'interno del mulino, sono solo nodi e legami. È interessante notare come, mentre in passato Hofstadter aveva caratterizzato l'alone concettuale come lo spazio delle possibili variazioni, ovvero come la "sfera controfattuale implicita", nel saggio del 2001 la forma dei concetti è *non-sferica*: «parole e concetti sono molto lontani dall'essere regioni convesse delimitate in maniera regolare nello spazio mentale; la polisemia (il possesso di molteplici significati) e la metafora rendono le regioni complesse e idiosincratiche» (*ivi*, p. 511).

Altro aspetto davvero rilevante è la rinnovata attenzione al linguaggio naturale, la cui trattazione è assente con le dovute spiegazioni e giustificazioni all'interno dei modelli cognitivi, ma che non può essere trascurata nell'elaborare una teoria dei concetti. Il linguaggio è visto come uno dei modi in cui è possibile riattivare i concetti unitamente agli stimoli percettivi. Perciò, termini e locuzioni del linguaggio costituiscono una sorta di attrattori da un punto di vista esternalista, ma anche uno dei modi in cui si accede al concetto nella sua interezza, e quindi si potrebbe dire uno dei nodi che ricadono nella nube di nodi costituiscono il concetto nella interezza. In questa ottica, il linguaggio diventa una sorta di indicatore che punta sui concetti, sia preso nei suoi termini singoli, sia per quanto riguarda descrizioni linguistiche più o meno lunghe di situazioni gradualmente più complesse. La comunicazione stessa fra individui che condividono la stessa lingua, o che sono in grado di comprendere la stessa lingua, è vista come un modo di creazione di analogie inter-mentali: «poiché ritengo che la metafora e l'analogia sono lo stesso fenomeno, ne consegue che io credo che tutta la comunicazione avviene per via analogica» (*ivi*, p. 526). Attraverso le etichette linguistiche un parlante riesce a evocare nell'ascoltatore, in senso statistico e con un certo grado di approssimazione, concetti analoghi a quelli che sta pensando e a cui il suo discorso si riferisce. Si parla anche in questo caso di analogie, perché ogni mente ha la propria struttura concettuale e il linguaggio è costituito di attrattori che evocano concetti (insiemi concettuali) che sono diversi da individuo a individuo.

Una questione potrebbe sorgere in merito alla funzione svolta dai termini che supportano le relazioni sintattiche all'interno delle proposizioni di cui si compongono le descrizioni linguistiche delle situazioni, quelli che Aristotele definiva "sincategorematici". Essi puntano a concetti specifici? La risposta potrebbe essere negativa, se si ritiene che essi indichino soltanto il modo in cui i concetti sono interrelati, e siano dunque, per così dire, termini che stanno per i processi di attivazione fra i concetti più che per i concetti stessi, e dunque termini la cui comprensione deve essere pensata come e non appresa ma innata in senso chomskiano, cioè come forme generative sintattiche fissate, all'interno di un repertorio predefinito, negli anni dell'apprendimento della lingua. Tuttavia, la risposta potrebbe anche essere positiva. I termini sincategorematici possono

puntare a concetti se vengono esplicitati, cioè se diviene oggetto dell'attenzione del pensiero la loro natura relazionale. Alla loro esplicitazione concorrono le metafore spaziali riprese dalla percezione visiva che sembrano essere connaturate al ragionamento umano. Verso una tale ipotesi si può dire che propenda Foundalis (2006) nell'implementare PHAEACO, ma, come abbiamo visto, l'idea di una pensiero permeato da metafore riprese dalla percezione dello spazio era già presente in Hofstadter (1979) e, inoltre, proprio allo scopo della sua implementazione sono presenti nella rete semantica di ogni modello nodi che esprimono relazioni di ordine ("predecessore", "successore") applicabili a ogni tipo di concetto, da quelli più concreti in diretta corrispondenza con le loro istanze specifiche a quelli più astratti. Tuttavia, il problema rimane aperto ad ulteriori precisazioni e investigazioni.

Altro aspetto non secondario della teoria è quello che riguarda i tempi della sua enunciazione. La teoria dei concetti come analogie viene, infatti, proposta da Hofstadter *in seguito* allo sviluppo di gran parte dei modelli subcognitivi, come se ne fosse una conseguenza diretta. In questo senso si può dire che la teoria nasca da una base sperimentale, simulativa, piuttosto ampia, la quale comprende la modellizzazione di processi di categorizzazione, di *retrieval* di esperienze memorizzate per via di somiglianza, di creazione di analogie, di mescolanza concettuale (*frame blending*), di concettualizzazione attraverso operazioni di composizione compiute su elementi misurati in modo statistico, la quale si traduce, appunto, nei concetti come nubi sfrangiate formate da concetti costituenti, tenendo presente che non c'è un vincolo specifico a che un concetto piuttosto che un altro sia costituente; è il contesto a determinarlo. In tal modo viene superata la rigidità delle pur flessibili tecniche di rappresentazione della conoscenza costituite dai *frame*. La proposta della teoria in seguito allo sviluppo dei modelli è, perciò, un punto a favore di una visione scientifica in senso proprio della ricerca simulativa. Essa ha anche il vantaggio di porre dei punti fermi in merito alle teorie proposte e di evidenziare la loro non circolarità, grazie appunto alle prestazioni positive messe in atto dai programmi che implementano i modelli.

Si può dire, tirando le somme del discorso, che la parte di *Gedankenexperiment* che all'inizio abbiamo provocatoriamente affermato essere presente negli esperimenti simulativi, ha qui il ruolo di generalizzazione a partire dai meccanismi delineati in maniera non ambigua e dai dati in quanto risultati della prestazione. Questa, tuttavia, è la parte che avrebbe in qualsiasi esperimento scientifico, quella dell'*invenzione* della teoria adatta a spiegarlo. Ciò provoca un distacco dallo svuotamento formale di cui vengono accusate a volte le strategie simulative dai sostenitori di una visione metafisica materialistica della realtà.

Infine, proseguendo in questa direzione si può anche pensare la teoria dei concetti come analogie alla stregua di quella generalizzazione che costituisce uno degli obiettivi dei modelli subcognitivi, pensati per agire in un microdominio, ma con l'intenzione di modellare capacità dominio-indipendenti, le quali tutte sono riconducibili al processo di creazione di analogie. Se, perciò, questa teoria mostra di fornire spiegazioni per molte delle proprietà che una teoria dei concetti deve poter

spiegare sia dal punto di vista psicologico che filosofico, se ne deduce un alto potere esplicativo, che tuttavia non sana tutti i problemi che la teoria, unitamente alla TCCL, mancano di spiegare, come ad esempio una più adeguata descrizione del modo in cui avvengono i processi di descrizione, o l'esatta natura delle primitive concettuali e relazionali dalle quali la teoria sembra dipendere per risolvere il problema del riferimento, o, ancora, una giustificazione più dettagliata del modo in cui il linguaggio naturale è trattato dalla teoria stessa, che pure si dimostra molto adatta a rendere conto dei meccanismi della traduzione inter-linguistica[17], grazie al suo corollario sulla "comunicazione analogica".

4.7 Conclusione ricorsiva

Nel corso di questo lavoro abbiamo cercato di dare conto di una spiegazione dei meccanismi mentali che si situa a un livello intermedio sia per quanto riguarda il modo di guardare, cioè di descrivere, il sistema mente-cervello, oggetto globale di indagine della scienze cognitive, sia in merito ai principi e alle strategie utilizzate dalle varie metodologie simulative per studiare i meccanismi di pensiero. Abbiamo visto come i modelli che ricadono in questo approccio non abbandonano un'impostazione funzionalista, né negano il ruolo centrale della rappresentazione, resistendo alle critiche portate a questa idea da parte del connessionismo purista, cioè subsimbolico, ed essendo in linea con ciò che proprio in questi tempi viene ribadito con forza, cioè che la presenza di un qualche meccanismo rappresentazionale nel sistema mente-cervello va necessariamente postulata per spiegare moltissimi fenomeni cognitivi, di contro alle posizioni espresse dagli eliminativisti. Tuttavia, questo non esime dal dover penetrare a fondo la nozione di rappresentazione e dal considerare quale sia il modo migliore di pensare a un sistema rappresentazionale, affinché la spiegazione del mentale non sia del tutto scollegata, ricreando una sorta di ostacolo dualistico, dal sistema che la implementa. Il riduzionismo si configura come una tensione positiva nel regolare la ricerca: «non si può parlare solo di neuroni per spiegare la mente, così come non si può parlare di geni per spiegare gli organismi. La mente non può essere ridotta al cervello. Tuttavia, forse un giorno, in qualche modo, lo sarà»[18].

Inoltre, si è visto come, per non cadere nel tranello posto dalla *regressio ad infinitum* tipica dei sistemi rappresentativi auto-giustificati, è necessario introdurre la nozione di *emergenza*, secondo la quale, sia la mente e l'azione intelligente umana sono pensate come risultato dell'interazione di molte micro-azioni più semplici, sia la conoscenza è frutto di un sistema relazionale gerarchico e dipendente dal contesto di concetti posti su più livelli, la cui composizione organizzata genera di volta in volta concetti di ordine superiore. Il livello simbolico è, dunque, in questa prospettiva

[17] Per un'analisi dettagliata dei problemi posti dalla traduzione e legati alla teoria dei concetti come analogie si rimanda a Hofstadter (1997).
[18] Hofstadter, comunicazione personale.

emergente da un livello di simbolicità inferiore, in una cascata di livelli che trova fine solo in accoppiamenti diretti funzione-meccanismo di natura estremamente semplice. In questo, e solo in questo, ha senso vedere un'analogia fra menti e calcolatori, cioè nel punto esatto di convergenza fra operazioni basilari della macchina (*hardware*) e operazioni non riducibili del *software*.

Riassumendo, l'approccio definito subcognitivo ha prodotto una serie di sistemi di simulazione della attività intelligente, che sono modelli:

- di una teoria del sistema mente-cervello che non trascuri l'uno e l'altro dei membri del sistema, procedendo all'individuazione di un livello di analisi intermedio del rapporto fra fenomeni mentali ed eventi cerebrali. Uno degli scopi principali è quello di individuare l'opportuno livello di analisi dei fenomeni mentali, non così distante ed eterogeneo rispetto al fenomeno da spiegare come quello proposto dal connessionismo (quello delle reti neurali, modello semplificato delle reti neuronali), ma allo stesso tempo alternativo anche alle tradizionale visione simbolico-rappresentativa della mente, in cui l'elaborazione è garantita dalla manipolazione formale di simboli fortemente centralizzata. Il risultato è che per simulare meccanismi del pensiero come la percezione di alto livello o la creazione di analogie occorre fissare una serie di *primitive relazionali*, cioè di concetti che permettono l'individuazione di relazioni di ordine ed equivalenza di classe, sulla base però di altri concetti, *di livello equivalente*, che costituiscono le classi o itipi, le cui istanze risiedono nello spazio percettivo. Ciò fa sì che i modelli siano anche modelli

- di una concezione distribuita, ma rappresentativa della conoscenza, in cui vengono esplicitati e posti in un'architettura funzionale i meccanismi processuali alla base della percezione di strutture, che *usualmente sfuggono* all'attenzione cosciente e il cui resoconto introspettivo è inevitabilmente viziato dalla riflessione a posteriori del soggetto, con il rischio che ciò che viene descritto non sia il processo, ma ciò che del processo viene ricordato, o, per meglio dire, esplicitato nel ricordo. Questo ha conseguenze anche sulla metodologia implicata nella valutazione dei risultati conseguiti da questi sistemi. La loro plausibilità psicologica è, per una buona parte, tanto maggiore quanto più grande è la coincidenza fra l'insieme delle diverse risposte prodotte dal modello con l'insieme delle risposte date da soggetti umani. Poiché il livello dei fenomeni mentali indagato è per definizione sottoposto, e dunque sfuggente, a quello della attenzione cosciente, non rimane che il raffronto dei risultati, anche a seguito di opportune variazioni nell'architettura dei modelli, per la valutazione dell'efficacia e del conseguimento degli obiettivi che ci si pone con la costruzione dei modelli. In altri termini, il resoconto introspettivo, sui cui si basarono i primi realizzatori di programmi di IA con l'esplicito fine di riprodurre sistemi psicologicamente plausibili, è negato (Newell, Simon, 1972). Nell'individuare il minimo

livello *ultra*-neurale del pensiero i modelli subcognitivi costituiscono ancora una via per l'indagine della rappresentazione intesa in senso simbolico, anche se non nel senso di simboli logico-formali. Perciò essi sono anche modelli

- di una teoria dei concetti che li considera in senso lato e a tutti i livelli di complessità, come *unità strutturate* (*pattern*) a partire da unità più semplici, ma *sempre e soltanto* sulla base di un contesto semantico-percettivo, costituito dall'insieme della rete semantica (che viene a essere un dominio semantico) e dal materiale presente nella memoria di lavoro. Le unità più semplici su cui tutto il processo si basa sono *concetti primitivi relazionali* comuni a tutti i modelli e che rappresentano, consentendo una generalizzazione dei vari fenomeni dominio-specifici, la parte teorica innata dei modelli. Il ragionamento analogico e la percezione di alto livello sono possibili soltanto a partire da alcuni concetti relazioni costanti che permettono l'organizzazione della situazione percepita, sotto l'influsso del bagaglio epistemico già posseduto, e che realizzano quella intuizione spaziale che guida la percezione di situazioni (oggetti, eventi, strutture sociali) del mondo reale. Tali primitive possono riscontrarsi in ogni modello. Il funzionamento di tali sistemi vuole essere una prova a conferma della loro effettiva presenza a un qualche livello della mente. A quale livello è, in definitiva, una questione aperta. Tuttavia, potrebbe essere questo tipo di concetti relazionali (identità di classe, successione) a essere indagato a livello cerebrale. Esperimenti di *neuroimaging* sui domini specifici di questi modelli sono stati tentati molto di recente (Geake, Hansen, 2005), con risultati conformi a questo tipo di ricerche, cioè l'individuazione di una specifica sottoarea cerebrale dell'area di Broca coinvolta nello svolgimento dei compiti analogici di COPYCAT e attiva preliminarmente all'elaborazione linguistica. La grana di questo tipo di sperimentazioni non è ancora fine a tal punto da poter individuare attività di maggior dettaglio. Se questa sia un'impossibilità di principio o contingente è argomento che sarà ancora, prevedibilmente, argomento di un lungo dibattito. Per ora, non rimane da aggiungere che i modelli subcognitivi sono, infine, modelli

- di una teoria che unifica la spiegazione dei concetti e quella del ragionamento analogico.

I problemi non risolti di tale teoria ancora in via di definizione non nascondono il suo richiamare, che si impone in modo quasi immediato, lo humeano io fascio di percezioni, concetto fra gli altri concetti, tutti, in ultima analisi, fasci di analogie:

> Noi non siamo altro che fasci o collezioni di differenti percezioni che si susseguono con una inconcepibile rapidità, in un perpetuo flusso e movimento. I nostri occhi non possono girare nelle loro orbite senza variare le nostre percezioni. Il nostro pensiero è ancora più variabile della nostra vista, e tutti

gli altri sensi e facoltà contribuiscono a questo cambiamento; né esiste forse un solo potere dell'anima che resti identico, senza alterazione, un momento. La mente è una specie di teatro, dove le diverse percezioni fanno la loro apparizione, passano e ripassano, scivolano e si mescolano con un'infinita varietà di atteggiamenti e di situazioni. Né c'è, propriamente, in essa nessuna semplicità in un dato tempo, né identità in tempi differenti, qualunque sia l'inclinazione naturale che abbiamo ad immaginare quella semplicità e identità. E non si fraintenda il paragone del teatro: a costituire la mente non c'è altro che le percezioni successive: noi non abbiamo la più lontana nozione del posto dove queste scene vengono rappresentate, o del materiale di cui è composta. (Hume, 1739-40/1971, pp. 264-265)

BIBLIOGRAFIA

- ABELSON R. P. (1968), «Simulation of Social Behavior», in G. Lindzey, E. Aronson (eds.), *Handbook of Social Psychology*, Addison-Wesley, Reading (Mass.), vol II, pp. 274-356.

- ACKLEY D. H., HINTON G. E., SEJNOWSKI T. J. (1985), «A Learning Algorithm for Boltzmann Machines», in J. A. Anderson, E. Rosenfeld (eds.), *Neurocomputing: Foundations of Research*, MIT Press, Cambridge, Mass., (1988).

- ANDERSON J. R., LEBIÈRE C. (1999), *The Atomic Components of Thought*, Erlbaum, Hillsdale, NJ.

- ATKINSON R. C., SHIFFRIN R. M. (1968), «Human memory: a Purposed system and its control process», in K. W. Spence, J. T. Spence (eds.), *The Psychology of Learning and Motivation*, Academic Press, New York, vol. 2, pp. 89-195.

- BADDELEY A. D. (1986), *Working Memory*, Oxford University Press, Oxford (trad. it. *La memoria di lavoro*, Raffallo Cortina, Milano, 1990).

- BARA B. G. (1978), «La validazione dei modelli di simulazione», in B. G. Bara (a cura di), *Intelligenza artificiale*, Franco Angeli, Milano, pp. 67-92.

- BARSALOU L. W. (1983), «Ad hoc categories», in *Memory and Cognition*, 11, pp. 211-217.

- BECKER J. D (1973), «A model for the encoding of experiential information», in R. C. Schank, K. M. Kolby (eds.), *Computer Models of Thought and Language*, Freeman, San Francisco, CA, pp. 396-435.

- BIEDERMAN I. (1987), «Recognition by components: A theory of human image understanding», in *Psychological Review*, 94, pp. 115-147.

- BODEN M. (1986), *Artificial Intelligence and Natural Man*, 2nd edition, MIT Press, Cambridge Mass. (trad. it. a cura di Maurizio Matteuzzi, *Intelligenza umana e intelligenza artificiale*, Tecniche Nuove, Milano, 1993).

- BODEN M. (1990), *The Creative Mind: Muths and Mechanisms*, Basic Books, New York.

- BODEN M. (ed.) (1994), *Dimensions of Creativity*, The MIT Press, Cambridge, Mass.

- BONGARD M. (1970), *Pattern Recognition*, Spartan Books, Rochelle Park, NJ.

- BORGHI A. M. (1996), *L'organizzazione della conoscenza. Aspetti e problemi*, Patron, Bologna.

- BORGHI A. M. (2002), «Concetti e azione», in A. M. Borghi, T. Iachini (a cura di), *Scienze della mente*, Il Mulino, Bologna, pp. 203-222.

- BRAITENBERG V. (1984), *Vehicles: Experiments in Synthetic Psychology*, MIT Press, Cambridge, Mass. (trad. it. a cura di Nicola Bruno e Lidia Martinuzzi, *I veicoli pensanti*, Garzanti, Milano, 1984).

- BRUNER S. J., GOODNOW J. J., AUSTIN G. A. (1956), *A study of thinking*, Wiley and sons, New York (trad. it. a cura di E. Riverso, *Il pensiero: strategie e categorie*, Armando, Roma, 1969).

- BURNSTEIN M. H. (1986), «Concept formation by incremental analogical reasoning and debugging», in R. S. Michalski, J. G. Carbonell, T. M. Mitchell (eds.), *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufmann, Los Altos, CA, pp. 351-370.

- CALABI C. (2005), «Spiegazione e riduzione: Leibniz e i filosofi della mente», in S. Gensini (a cura di), *Linguaggio, mente, conoscenza. Intorno a Leibniz*, Carocci, Roma, pp. 193-214.

- CARBONELL J. G. (1983), «Learning by analogy: Formulating and generalizing plans from past experience», in R. S. Michalski, J. G. Carbonell, T. M. Mitchell (eds.), *Machine Learning: An Artificial Intelligence Approach*, Tioga, Palo Alto, CA, pp.136-162.

- CHALMERS D. J. (1996), *The Conscious Mind*, Oxford University Press, Oxford (trad. it. a cura di Alfredo Paternoster e Cristina Meini, *La mente cosciente*, McGraw-Hill, Milano, 1999).

- CHALMERS D. J., FRENCH R. M., HOFSTADTER D. R. (1992), «High-level perception, representation, and analogy: A critique of artificial intelligence metodology», in *Journal of Experimental and Theoretical Artificial Intelligence*, 4, pp. 185-211 (trad. it. in Hofstadter & FARG (1995), pp. 187-212).

- CHURCHLAND P. M. (1995), *The Engine of Reason, the Seat of the Soul: A Philosophical Journey into the Brain*, MIT Press, Cambridge (MA) (trad. it. a cura di Pier Daniele Napolitani, *Il motore della ragione, la sede dell'anima*, Il saggiatore, Milano, 1998).

- CLOWES M. B. (1971), «On seeing things», in *Artificial Intelligence*, 2, pp. 79-116.

- COLBY K. M. (1963), «Computer Simulation of Neurotic Process», in S. S. Tomkins, S. Messick (eds.), *Computer Simulation of Personality: Frontier of Psychological Research*, Wiley, New York, pp. 165-180.

- COLIVA A. (2004), *I concetti. Teorie ed esercizi*, Carocci, Roma.

- CORDESCHI R. (2002), *The Discovery of the Artificial: Behavior, Mind and Machines Before and Beyond Cybernetics*, Kluwer Academic Publishers, Dordrecht (trad. ingl. ampliata di *La scoperta dell'artificiale*, Masson/Zanichelli, Milano/Bologna, 1998).

- CORDESCHI R., FRIXIONE M. (2006), «Computazionalismo sotto attacco», in P. Cherubini, P. Giaretta, M. Marraffa, A. Paternoster (a cura di), *Cognizione e computazione. Problemi, metodi e prospettive delle spiegazioni computazionali nelle scienze cognitive*, CLEUP, Padova.

- CRAIK K. J. W. (1943), *The Nature of Explanation*, Cambridge University Press, Cambridge.

- CULLINGFORD R. E. (1978), *Script application: Computer understanding of newspaper stories*, Tech. Rep. 116, Yale University, Department of Computer Science, Ph.D. thesis.

- CUMMINS R. (1989), *Meaning and mental representation*, MIT Press, Cambridge, Mass.

- DARTNALL T. (ed.) (2002), *Creativity, cognition, and knowledge: an interaction*, CT, Westport.

- DENIS M., MELLET E., KOSSLYN S. M (EDS.) (2004), *Neuroimaging of mental imagery*, Psychology Press, Hove.

- DENNETT D. C. (1978), *Brainstorms: philosophical essays on mind and psychology*, Harvester Press, Hassocks (trad. it. a cura di Lauro Colasanti, *Brainstorms: saggi filosofici sulla mente e la psicologia*, Adelphi, Milano, 1991).

- DENNETT D. C. (1980), «Il latte dell'intenzionalità umana», in Searle (1980/1984), pp. 94-100.

- Dennett D. C. (1982), «How to study consciousness empirically: or nothing comes to mind», in *Synthese*, 53, pp. 159-180)

- DENNETT D. C. (1989), *The intentional stance*, The MIT Press, London (trad. it. a cura di Erica Bassato, *L'atteggiamento intenzionale*, Il Mulino, Bologna, 1993).

- DENNETT D. C. (1991), *Consciousness explained*, Little Brown, Boston (trad. it. a cura di Lauro Colasanti, *Coscienza*, Rizzoli, Milano, 1993).

- DENNETT D. C. (1998), «Il mito della doppia trasduzione», in *Atque*, 16. pp. 11-26.

- DENNETT D. C. (2005), *Sweet Dreams. Philosophical Obstacles to a Science of Consciousness*, MIT Press, Cambridge, Mass. (trad. it. a cura di Antonino Cilluffo, *Sweet Dreams. Illusioni filosofiche sulla coscienza*, Raffaello Cortina, Milano, 2006).

- DIETTERICH T. G, MICHALSKI R. S. (1985), «Discovering patterns in sequences of events», in *Artificial Intelligence*, 25, 1985, pp. 187-232.

- DREYFUS H. L. (1981), «From micro-world to knowledge representation: A.I. at an impasse», in J. Haugeland, *Mind Design: Philosophy, Psychology, Artificial Intelligence*, MIT Press, Cambridge, Mass., pp. 161-204 (trad. it. *Progettare la mente: filosofia, psicologia, intelligenza artificiale*, Il Mulino, Bologna, 1989, pp. 177-219).

- EIBEN A. E., RUDOLPH G. (1999), «Theory of evolutionary algorithms: a bird's eye view», in *Theoretical Computer Science*, 229, pp. 3-9.

- ERNST G., NEWELL A. (1969), *GPS: A Case Study in Generality and Problem Solving*, Academic Press, New York.

- EVANS T. G. (1968), «A program for the solution of a class of geometric analogy intelligence questions», in M. Minsky (ed.), *Semantic Information Processing*, MIT Press, Cambridge, Mass., 1968, pp. 272-277.

- FALKENHAINER B., FORBUS K. D., GENTNER D. (1989), «The structure-mapping engine: Algorithm and examples», in *Artificial Intelligence*, 41, pp. 1-63.

- FAUCONNIER G., TURNER M. (2002), *The way we think: conceptual blending and the mind's hidden complexities*, Basic Books, New York.

- FISHER SERVI G. (2001), *Quando l'eccezione è la regola. Le logiche non monotone*, McGraw-Hill, Milano.

- FLOREANO D., MATTIUSSI C. (2002), *Manuale sulle reti neurali*, Il Mulino, Bologna.

- FODOR J. A. (1976), *The Language of Thought*, Harvester Press, Hassocks.

- FORBUS K., GENTNER D., LAW K. (1995), «MAC/FAC: A Model of Similarity-Based Retrival», in *Cognitive Science*, 19, pp.141-205.

- FOUNDALIS H. E. (2006), *Phaeaco: A Cognitive Architecture Inspired by Bongard's Problems*, Ph.D. Dissertation, Indiana University, Bloomington, IN.

- FRANCHI S. (2004), «Teoria dei giochi e intelligenza artificiale», in F. Bianchini, M. Matteuzzi, *Percezione linguaggio coscienza. Percorsi tra cognizione e intelligenza artificiale, Discipline Filosofiche*, 2, Quodlibet, Macerata, pp. 63-88.

- FRENCH R. M. (1990), «Subcognition and the Limits of the Turing Test», in *Mind*, 99, pp. 53-65).

- FRENCH R. M. (1995), *The Subtlety of Sameness*, MIT Press, Cambridge, Mass.

- FRENCH R. M.(in corso di pubblicazione), «The dynamics of the computational modeling of analogy-making», in P. Fishwick (ed.), *CRC Handbook of Dynamic Systems Modeling*, CRC Press LLC, Boca Raton, Fl.

- FRENCH R. M., HOFSTADTER, D. R. (1991), «Tabletop: An Emergent, Stochastic Model of Analogy-Making», in *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, Hillsdale, NJ, pp. 708-713.

- GALLUP G. (1970), «Chimpanzees: Self-Recognition», in *Science*, 167, pp. 86-87.

- GEAKE J. G., HANSEN P. C. (2005), «Neural correlates of intelligenze as revealed by fMRI of fluid analogies», in *NeuroImage*, 26,2, pp. 555-564.

- GENTNER D. (1983), «Structure-Mapping: A Theoretical Framework for Analogy», in *Cognitive Science*, 7[2], pp. 155-170.

- GOLDBERG H. G., REDDY D. R., SUSLICK R. L. (1974), «Parameter-indipendent machine segmentation and labeling», in *Proceedings of IEEE Symposium Speech Recognition*, Carnegie-Mellon University, Pittsburgh, Pa, pp. 106-111.

- GOODMAN N. (1972), «Seven Strictures on Similarity», in Id., *Problem and Projects*, Bobbs-Merril Company, Indianapolis, Ind, and New York, pp. 437-447.

- GOPNIK A., MELTZOFF A. (1997), *Words, Thoughts and Theories*, MIT Press, Cambridge, Mass.

- GOZZANO S. (1997), *Storia e teorie dell'intenzionalità*, Laterza, Roma.

- GREBERT I., STORK D., KEESING R., MINS S. (1991), «Network generalization for production: Learning and producing styled letterforms», in *Proceedings of the Neural Information Processing Systems Conference*, pp. 1118-1124.

- GREBERT I., STORK D., KEESING R., MINS S. (1992), «Connectionist generalization for production: An example from GridFont», in *Neural Networks*, 5, pp. 699-710.

- GREINER R. (1988), «Abstraction-based analogical inference», in D. H. Helman (ed.), *Analogical Reasoning: Perspectives of Artificial Intelligence, Cognitive Science and Philosophy*, Kluwer Academics Publishers, Dordrecht, pp. 147-170.

- GUNDERSON K. (1964), «The Imitation Game», in A. Anderson (ed.), *Minds and Machines*, Prentice-Hall, Englewood Cliffs (NJ), pp. 60-71.

- GUZMAN A. (1968), «Decomposition of a visual scene into three-dimensional bodies», in *American Federation of Information Processing Societies Fall Joint Conferences*, 33, pp.291-304.

- HALL R. P. (1989), «Computational Approaches to Analogical Reasoning: A Comparative Analysis», in *Artificial Intelligence*, 39, pp. 39-120.

- HAMPTON J. (1988), «Overextension of conjunctive concepts: evidence for a unitary model of concepts tipicality and class inclusion», in *Journal of Experimental Psychology: Learning Memory and Cognition*, 14, pp. 12-32.

- HARNAD S. (1990), «The Symbol Grounding Problem», in *Physica D*, 42, pp. 335-346.

- HARNAD S. (2003), «The Symbol Grounding Problem», in L. Nadel (editor-in-chief), *Encyclopedia of Cgnitive Science*, Nature Publishing Group, London.

- HAUGELAND J. (1980), «Programmi, poteri causali e intenzionalità», in Searle (1980/1984), pp. 107-113.

- HAUGELAND J. (1981), *Mind Design. Philosophy, Psychology, Artificial Intelligence*, MIT Press, Cambridge, Mass. (trad. it. di Paola Amaldi e Simone Gozzano, *Progettare la mente. Filosofia, psicologia intelligenza artificiale*, Il Mulino, Bologna, 1989).

- HEBB D. O. (1949), *The organization of behavior: a neuropsychological theory*, John Wiley, New York (trad. it. *L'organizzazione del comportamento*, Franco Angeli, Milano, 1975).

- HEWITT C. (1977), «Viewing control structures as patterns of passing messagge», in *Journal of Artificial Intelligence*, 8-3, pp. 323-364.

- HINTON G. E., SEJNOWSKI T. J. (1986), «Learning and Relearning in Boltzmann Machines», in D. E Rumelhart, J. L. McClelland *et.al.*, *Parallel Distribuited Processing: Explorations in teh Microstructure of Cognition*, MIT Press, Cambridge, Mass., 1986, vol. I, cap. 7.

- HOFSTADTER D. R. (1979), *Gödel, Escher, Bach: an Eternal Golden Braid*, New York, Basic Books (ediz. it. a cura di Giuseppe Tratteur, *Gödel, Escher, Bach: un'Eterna Ghirlanda Brillante*, Milano, Adelphi, 1984).

- HOFSTADTER D. R. (1981), «How might analogy, the core of human thinking, be understood by computers?», in *Scientific American*, 245, pp. 18-30 (trad. it. «Come possono i calcolatori comprendere l'analogia, il nucleo del pensiero umano», in *Le Scienze*, 159, novembre 1981, pp. 140-148).

- HOFSTADTER D. R. (1982), *SEEK-WHENCE: A Project in Pattern Undestanding*, tech. report n. 3, Center for Research on Concepts and Cognition, Indiana University, Bloomington, IN.

- HOFSTADTER D. R. (1983a), «The Architecture of Jumbo», in R. Michalski, J. Carbonell e T. Mitchell (eds), *Proceedings of the International Machine Learning Workshop*, University of Illinois, Urbana, Ill, pp. 161-170 (trad. it. e versione ampliata nella trad. it. di Hofstadter & FARG (1995), «L'architettura di Jumbo», pp. 111-142).

- HOFSTADTER D. R. (1983b), *On Seeking Whence*, tech. report n. 5, Center for Research on Concepts and Cognition, Indiana University, Bloomington, IN.

- HOFSTADTER D. R. (1985a), *Methamagical Themas: Questing for the Essence of Mind and Pattern*, Basic Books, New York.

- HOFSTADTER, D. R. (1985b), «On the Seeming Paradox of Mechanizing Creativity», in D. R. Hofstadter (1985a), pp. 526-546 (trad. it. parziale «Si può meccanizzare la creatività», in *Le scienze*, 171, 1982, pp. 164-174).

- HOFSTADTER D. R. (1985c), «Waking up from the Boolean Dream, or, Subcognition as Computation», in D. R. Hofstadter (1985a), pp. 631-665.

- HOFSTADTER D. R. (1985d), «Variation on a Theme as tha Crux of Creativity», in D. R. Hofstadter (1985a), pp. 232-259.

- HOFSTADTER D. R. (1985e), «Analogies and Roles in Human and Machine Thinking», in D. R. Hofstadter (1985a), pp. 547-603.

- HOFSTADTER D. R. (1985f), «Waking Up from the Boolean Dream, or, Subcognition as Computation», in D. R. Hofstadter (1985a), pp. 631-665.

- HOFSTADTER D. R. (1988), «Common Sense and Conceptual Halos», in *Behavioral and Brain Sciences*, 11, 1, pp. 35-37 (trad. it. nella traduzione italiana di Smolensky (1988), *Senso comune e aloni concettuali*», pp. 157-160).

- HOFSTADTER D. R. (1991), *A Short Compendium of Me-Too's and Related Phenomena: Mental Fluidity as Revealed in Everyday Conversation*, CRCC Technical Report n. 57. Center for Reasearch on Concepts and Cognition, Indiana University, Bloomington, Ind.

- HOFSTADTER D. R. (1994), «How could a Copycat ever be creative?», in T. Dartnall (ed.), *Artificial Intelligence and Creativity: An Interdiciplinary Approach*, Kluwer Academic Publischers, The Netherlands, pp. 405-424.

- HOFTSDATER D. R. (1995a), «To Seek Whence Cometh a Sequence», in Hofstadter & FARG (1995), pp. 13-86 (trad. it. «Successioni: un successone», nella trad. it. di Hofstadter & FARG (1995), pp. 27-100).

- HOFSTDATER D. R. (1995b), «Prolegomena to Any Future Metacat», in Hofstadter & FARG (1995), pp. 307-318 (trad. it. «Prolegomeni ad ogni futuro Metacat», nella trad. it. di Hofstadter & FARG (1995), pp. 331-342).

- HOFSTADTER D. R. (1995c), «The Knotty Problem of Evaluating Research in AI and Cognitive Science», Hofstadter & FARG (1995), pp. 359-376 (trad. it. «Lo spinoso problema di valutare la ricerca in IA e nella scienza cognitiva», nella trad. it. di Hofstadter & FARG (1995), pp. 385-404).

- HOFSTADTER D. R. (1995d), «Epilogue: On Computers, Creativity. Credit, Brain Mechanism and the Turing Test», in Hofstadter & FARG (1995), pp. 467-491 (trad. it. «Calcolatori, creatività, attribuzioni, meccanismi del cervello e test di Turing», nella trad. it. di Hofstadter & FARG (1995), pp. 499-524).

- HOFSTADTER D. R. (1995e), «On seeing A's and seeing As», in S. Franchi, G. Güzeldere, *Constructions of the Mind*, special issue of *Stanford Humanities Review*, 4, 2 (trad. it. a cura di Luigi Stringa, «Come vedere ha a che vedere con vedere come», in F. Bianchini, M. Matteuzzi, *Percezione linguaggio coscienza. Percorsi tra cognizione e intelligenza artificiale*, volume monografico di *Discipline Filosofiche*, Quodlibet, Macerata, 2, 2004, pp. 15-25).

- HOFSTADTER D. R. (1996), «Analogy-Making, Fluid Concepts, and Brain Mechanisms», in P. J. R. Millican, A. Clark (eds.), *The Legacy of Alan Turing. Vol. 2: Connectionism, Concepts and Folk Psychology*, Oxford University Press, Oxford.****

- HOFSTADTER D. R. (1997), *Le Ton beau de Marot*, Basic Books, New York.

- HOFSTADTER D. R. (2001), «Analogy as the core of cognition», in D. Gentner, K. J. Holyoak, B. N. Kokinov (eds.), *The analogical mind. Perspective from cognitive science*, MIT Press, Cambridge, Mass., pp. 499-538.

- HOFSTADTER D. R., CLOSSMAN G. A., MEREDITH M. J. (1980), *Shakespeare's Plays Weren't Written by him, but by Someone Else of the Same Name: An Essay on Intensionality and Frame-Based Knowledge-representation Systems*, tech. report n. 96, Computer Science Department, Indiana University, Bloomington, IN.

- HOFSTADTER D. R., DENNETT D. C. (1981), *The Mind's I. Fantasies and Reflections on Self and Soul*, Basci Books, New York (trad. it. a cura di Giuseppe Longo, *L'io della mente. Fantasie e riflessioni sul sé e sull'anima*, Adelphi, Milano, 1985).

- HOFSTADTER D. R. & THE FLUID ANALOGIES RESEARCH GROUP (FARG) (1995), *Fluid Concepts and Creative Analogies: Computer Models of the Foundamental Mechanisms of Thought*, Basic Books, New York (trad. it. a cura di Massimo Corbò, Isabella Giberti, Maurizio Codogno, *Concetti fluidi e analogie creative. Modelli per calcolatore dei meccanismi fondamentali del pensiero*, Adelphi, Milano, 1996).

- HOFSTADTER D. R., FRENCH R. M. (1992), «Probing the Emergent Behavior of Tabletop, an Architecture Uniting High-level Perception with Analogy-making», in *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, Hillsdale, NJ, pp. 528-533.

- HOFSTADTER D. R., FRENCH R. M. (1995a), «Tabletop, BattleOp, Ob-Platte, Potelbat, Belpatto, Platobet», in Hofstadter & FARG (1995), pp. 323-358 (trad. it. nella trad. it di Hofstadter & FARG (1995), «Tabletop, Battleop, Ob-platte, Potelbat, Belpatto, Platobet», pp. 347-384).

- HOFSTADTER D. R., FRENCH R. M. (1995b), «The Emergent Personality of Tabletop, a Perception-based Model of Analogy-making», in Hofstadter & FARG (1995), pp. 377-399 (trad. it. nella trad. it di Hofstadter & FARG (1995), «La personalità emergente di Tabletop: un modello del fare analogie basato sulla percezione», pp. 405-430).

- HOFSTADTER D. R., MARSHALL, J. (1998), «Making sense of analogies in Metacat», in K. Holyoak, D. Gentner, B. Kokinov (1998), pp. 118-123.

- HOFSTADTER D. R., MCGRAW G. (1995), «Letter Spirit: Esthetic Perception and Creative Play in the Rich Microcosm of the Roman Alphabet», in Hofstadter & FARG (1995), pp. 407-466 (trad. it. nella trad. it di Hofstadter & FARG (1995), «Letter Spirit: percezione estetica e gioco creativo nel ricco microcosmo dell'alfabeto latino», pp. 437-498).

- HOFSTADTER D. R., MITCHELL M. (1988), «Concepts, Analogies, and Creativity», in R. Goebel (ed.), *Proceedings of the Seventh Biennal Conference of the Canadian Society for Computationl Studies of Intelligence*, University of Alberta, Edmonton, pp. 94-101.

- HOLYOAK K., GENTNER D., KOKINOV B. (EDS.) (1998), *Advances in Analogy Research: Integration of Theory and Data form the Cognitive, Computaional and Neural Sciences*, New Bulgarian University, Sofia.

- HOLYOAK K., THAGARD P. (1989), «Analogical Mapping by Constraint Satisfaction», in *Cognitive Science*, 13, pp. 295-355.

- HUME D. (1739-40/1971), *Trattato sulla natura umana*, in *id*., *Opere*, a cura di E. Lecaldano, E. Mistretta, Laterza, Bari, pp. 1-665.

- HUMMEL J., HOLYOAK K. (1997), «Distribuited Representation of Structure: A Theory of Analogical Access and Mapping», in *Psychological Review*, 104, pp. 427-466.

- JOHNSON-LAIRD P. N. (1993), *Human and machine thinking*, LEA, Hillsdale, NJ (trad. it. a cura di Maurizio Riccucci, *Deduzione, induzione, creatività : pensiero umano e pensiero meccanico*, Il Mulino, Bologna, 1994).

- KANDEL E. R. (2006), *In Search of Memory. The Emergence of a New Science of Mind*, Northon & Company, New York (trad. it. a cura di Giuliana Olivero, *Alla ricerca della memoria. La storia di una nuova scienza della mente*, Codice edizioni Torino, 2007).

- KANERVA P. (1988), *Sparse Distribuited Memory*, MIT Press, Cambridge, Mass.

- KANT I. (1781-1787/1992), *Critica della ragion pura*, a cura di P. Chiodi, UTET, Torino.

- KAPLAN S., WEAVER M., FRENCH R. (1990), «Active Symbols and Internal Models: Towards a Cognitive Connectionism», in *AI & Society*, 4, pp. 51-71.

- KOKINOV B. (1994), «A Hybrid Model of Analogical Reasoning», in K. Holyoak, J. Barnden (eds.), *Advances in Connectionist and Neural Computation Theory. Vol 2: Analogica Connection*, Ablex Corporation, Norwood, NJ, pp. 247-318.

- KOKINOV B., FRENCH R. M. (2003), «Computational Models of Analogy-Making», in L. Nadel (ed.), *Encyclopedia of Cognitive Science*, Nature Publishing Group, London, vol. 1, pp. 113-118.

- KOKINOV B., PETROV A. (2001), «Integration of Memory and Reasoning in Analogy-Making: The AMBR Model», in D. Gentner, K. Holyoak, B. Kokinov (eds.), *The Analogical Mind. Perspective from Cognitive Science*, MIT Press, Cambridge, Mass., pp. 59-124.

- KOLODNER J. L. (1981), «Organization and retrieval in a conceptual memory for events», in *Proceedings of the Seventh International Joint Conference on Artificial Intelligence: IJCAI 81*, Morgan Kaufmann, Los Altos, Ca.

- KOLODNER J. L., SIMPSON R. L., SYCARA-CYRANSKI K. (1985), «A process model of case-based reasoning in problem solving», in *Proceeding IJCAI-85*, Los Angeles, CA, pp. 284-290.

- KOSSLYN S. M. (1980), *Image and mind*, Harward University Press, Cambridge, Mass.

- KOSSLYN S. M. (1983), *Ghosts in the minds machine: creating and using images in the brain*, Norton & Company, New York (trad. it. a cura di Gabriele Noferi, *Le immagini nella mente: creare e utilizzare immagini nel cervello*, Giunti, Firenze, 1989).

- KOSSLYN S. M. (1994), *Image and brain : the resolution of the imagery debate*, The MIT Press, Cambridge, Mass.

- KOTOVSKY K., SIMON H. A. (1973), «Empirical tests of a theory of human acquisition of concepts for sequential patterns», in *Cognitive Psychology*, 4, pp. 399-424.

- LAIRD J. E., NEWELL A., ROSENBLOOM P. S. (1987), «SOAR: an architecture for general intelligence», in *Artificial Intelligence*, 33, pp. 1-64.

- LARSON S. (1993), «Modeling Melodic Expectation: Using Three "Musical Forces" to Predict Melodic Continuations», in *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, Hillsdale, NJ, pp. 629-634 (tech. report n. 70, Center for Research on Concepts and Cognition, Indiana University, Bloomington, IN).

- LARSON S. (1997), «Seek Well: A Domain for Studying Melodic Expectation», in *Proceedings of the Joint International Conference: Fourth International Symposium on Systematic and Comparative Musicology and Second International Conference on Cognitive Musicology*, College of Europe at Brugge, Belgium, pp. 144-151 (tech. report n. 110, Center for Research on Concepts and Cognition, Indiana University, Bloomington, IN).

- LEBOWITZ M. (1980), *Generalization and memory in an integrated understandong system*, tech. rep. 186, Yale University, Department of Computer Science, Ph. D. thesis.

- LEGRENZI P. (1999), *Storia della psicologia*, Il Mulino, Bologna.

- Leibniz G. W (1705/1982), *Nuovi saggi sull'intelletto umano*, a cura di Massimo Mugnai, Editori Riuniti, Roma.

- LEIBNIZ G. W. (1710/2000), *Saggi di Teodicea*, in *id.*, *Scritti filosofici*, a cura di M. Mugnai, E. Pasini, UTET, Torino, vol. III, pp. 19-428.

- LEIBNIZ G. W. (1714/2001), *Monadologia*, a cura di Salvatore Cariati, Bompiani, Milano.

- LEIBNIZ G. W. (1963), *Saggi filosofici e lettere*, a cura di V. Mathieu, Laterza, Bari.

- LERDAHL F. (2001), *Tonal pitch space*, Oxford University Press, Oxford.

- LERDAHL F., JACKENDOFF R. (1983), *A generative theory of tonal music*, MIT Press, Cambridge, Mass.

- LESSER V. R., FENNELL R. D., ERMAN L. D., REDDY D. R. (1975), «Organization of the HEARSAY II Speech Understanding System», in *IEEE Transactions on Acoustics, Speech and Signal Processing*, 23, pp. 11-24.

- LINHARES A. (2000), «A glimpse at the metaphysics of Bongard problems», in *Artificial Intelligence*, 121, pp. 251-270.

- LINHARES A. (2005), «An active symbols theory of chess intuition», in *Minds and Machines*, 15, pp. 131-181.

- LOLLI G. (1994), *Introduzione*, in Turing (1992), pp. 7-23.

- LUCAS J. R. (1961), «Minds, Machines and Gödel», in *Philosophy*, 36, pp. 112-127.

- LUCCIO R. (1998), *Psicologia generale. Le frontiere della ricerca*, Laterza, Roma-Bari.

- MANZOTTO R., TAGLIASCO V. (2006), «Libertà e coscienza: un approccio basato sul processo», in *Sistemi intelligenti*, XVIII, 2, pp. 259-281.

- MARGULIS E. (2005), «A Model of Melodic Expectation», in *Music Perception*, 22, pp. 663-714.

- MARR D. (1982), *Vision; A Computational Investigation into the Human Representation and Processing of Visual Information*, W. H. Freeman, San Francisco.

- MARSHALL J. (1999), *Metacat: A Self-Watching Cognitive Architecture for Analogy-Making and High-Level Perception*, Ph.D. Dissertation, Indiana University, Bloomington, IN.

- MARSHALL J. (2002), «Metacat: a self-watching cognitive architecture for analogy-making», in W. D. Gray, C. D. Schunn (eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 631-636.

- MARSHALL J. (2006), «A self-watching model of analogy-making and perception», in *Journal of Experimental and Theoretical Artificial Intelligence*, 18(3), pp. 267-307.

- MATTEUZZI M. (1995), «Why AI is not a science», in *Constructions of the Mind Artificial Intelligence and the Humanities*, vol. mon. della rivista elettronica *Stanford Humanities Review*, 4,2 (trad. it. «Perché l'IA non è una scienza?», in *Discipline Filosofiche*, 6, 1996, pp. 233- 248).

- MCCARTHY J., MINSKY M. L., ROCHESTER N., SHANNON C. E. (1955), «A proposal for the Dartmouth Summer Research Project on Artificiail Intelligence», reperibile on line al sito http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html (trad. it. a cura di Gianlunca Paronitti, «Proposta di un progetto di ricerca estivo sull'intelligenza artificiale presso il Dartmouth College», in *Sistemi Intelligenti*, XVIII, 3, 2006, pp. 413-428).

- MCGRAW G. E. (1992), *Letter Spirit: Recognition and Creation of Letterforms Based on Fluid Concepts*, tech. report n. 61, Center for Research on Concepts and Cognition, Indiana University, Bloomington, IN.

- MCGRAW G. E. (1995), *Letter Spirit (part one): Emergent High-Level Perception of Letters Using Fluid Concepts*, Ph.D. Dissertation, Indiana University, Bloomington (IN).

- MCGRAW G. E., DRASIN D. (1993), «Recognition of Gridletters: Probing the Behavior of Three Competing Models», in T. E. Ahlswede (ed.), *Proceedings of the Fifth Midwest AI and Cognitive Science Society Conference*, Southern Illinois University, Carbondale IL, pp. 63-67.

- MCGRAW G. E., HOFSTADTER D. R. (2002), « Perception and Creation of Diverse Alphabetic Style», in T. Dartnall (ed.), *Creativity, Cognition and Knowledge: An Interaction*, Praeger, Westport, CT.

- MCGRAW G. E., HOFSTADTER D. R. (1993), «Letter Spirit: An Architecture for Creativity in a Micro-domain», in P. Torasso (ed.) *Advances in Artificial Intelligence, Third Congress of the Italian Association for Artificial Intelligence*, Torino, pp. 65-70.

- MCGRAW G. E., HOFSTADTER D. R. (1996), *Emergent Letter Perception: Implementing the Role Hypothesis*, tech. report n. 103, Center for Research on Concepts and Cognition, Indiana University, Bloomington, IN.

- MCGRAW G. E., REHLING J. A., GOLDSTONE R. (1994a), *Roles in Letter Perception: Human data and computer models*, tech. report n. 90, Center for Research on Concepts and Cognition, Indiana University, Bloomington (IN).

- MCGRAW G. E., REHLING J. A., GOLDSTONE R. (1994b), «Letter Perception: Toward a conceptual approach», in A. Ram, K. Eiselt (eds.), *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Erlbaum, Hillsdale, NJ, pp. 613-618.

- MEDIN D. L., SCHAFFER M. M. (1978), «Context theory of classification learning», in *Psychological Review*, 85, pp. 207-238.

- MEHLER J., DUPUOX E. (1990), *Naitre humain*, Jacob, Paris (trad. it. cura di Elena Mohlo, *Che cosa vede, sente, capisce un bambino sin dai primi giorni di vita*, Mondadori, Milano, 1992).

- MEINI C., PATERNOSTER A. (in corso di pubblicazione), «Categorization and Concepts: A Methodological Framework», in M. De Caro, F. Ferretti, M. Marraffa (eds.), *Cartographies of the Mind*, Kluwer, Dordrecht.

- MELANDRI E. (2004), *La linea e il circolo. Studio logico-filosofico sull'analogia*, Quodlibet, Macerata.

- MEREDITH M. J. (1986), *Seek- Whence: A Model in Pattern Perception*, tech. report n. 214, Computer Science Department, Indiana University, Bloomington (IN).

- MILLER G. A., JOHNSON-LAIRD P. N. (1976), *Language and perception*, Cambridge University Press, Cambridge, Mass.

- MINSKY M. (1966), «Artificial Intelligence», in *Scientific American*, 215, pp. 246-263.

- MINSKY M. (1975), «A Framework for Representing Knowledge», in P. H. Winston (ed.), *The Psycology of computer vision*, McGraw-Hill, New York, pp. 211-280 (trad. it. «Un sistema per la rappresentazione della conoscenza», in Haugeland (1981), pp. 107-142).

- MINSKY M. (1986), *The society of mind*, Simon and Schuster, New York (trad. it. a cura di Giuseppe Longo, *La società della mente*, Adelphi, Milano, 1989).

- MITCHELL M. (1993), *Analogy-Making as Perception*, MIT Press, Cambridge, Mass.

- MITCHELL M. (2001), «Analogy-Making as a Complex Adaptive System», in L. A. Segel, I. R. Cohen (eds.), *Design Principles for the Immune System and Other Distribuited Autonomous Systems*, Oxford University Press, New York.

- MITCHELL M. (2005), «Self-awareness and control in decentralized systems», in *Working Papers of the AAAI 2005 Spring Symposium on Metacognition in Computation*, AAAI Press, Menlo Park, Ca.

- MITCHELL M., HOFSTADTER D. R. (1990), «The emergence of understanding in a computer model of concepts and analogy-making», in *Physica D*, 42, pp. 322-334.

- MITCHELL M., HOFSTADTER D. R. (1994), «The Copycat Project: A Model of Mental Fluidity and Analogy-Making», in K. Holyoak, J. Barnden (eds.), *Advances in Connectionist and Neural Computation Theory. Vol 2: Analogical Connections*, Ablex Corporation, Norwood, NJ, pp. 31-112 (trad. it. in Hofstadter & FARG (1995), «Il progetto Copycat: un modello della fluidità mentale e della creazione di analogie» e «Panoramica su Copycat: paragone con lavori precedenti», pp. 225-290 e 297-322).

- MURPHY G. L., MEDIN D. L. (1985), «The role of theories in conceptual coherence», in *Psycological Review*, 92, pp. 289-316.

- MURPHY G. L. (2002), *The Big Book of Concepts*, MIT Press, Cmabridge, Mass.

- NANARD M., NANARD J., GANDARA M., PORTE N. (1989), «A Declarative approach for font design by incremental learning», in J. Andre, R. Hersch (eds.), *Raster Imaging and Digital Typography*, Cambridge University Press, Cambridge, pp. 71-82.

- NARMOUR E. (1992), *The analysis and cognition of melodic complexity. The implication-realization model*, University of Chicago Press, Chicago.

- NEWELL A. (1990), *Unifies Theories of Cognition*, Harward University Press, Cambridge, Mass.

- NEWELL A., BARTNETT J., FORGIE J., GREEN C., KLATT D., LICKLIDER J. C. L., MUNSON J., REDDY R., WOODS W. (1973), *Speech Understanding System: Final Report of a Study Group*, Elsevier/North-Holland, Amsterdam.

- NEWELL A., SHAW J.C., SIMON H.A. (1960), «Report on a general problem-solving program», *Proceedings of the International Conference on Information Processing [UNESCO House, Paris, France, June 13-23, 1959*], pp. 256-264.

- NEWELL A., SIMON H. A. (1972), *Human problem solving*, Prentice-Hall, Englewood Cliffs (NJ).

- NOSOFSKY R. M. (1988), «Exemplar-based accounts of relations between classification, recognition, and typicality», in *Journal of Experimental Psychology: Learning, Memory and Cognition*, 14, pp. 700-708.

- NOSOFSKY R. M., PALMERI T. J. (1997), «An exemplar-based random walk model of speeded categorization», in *Psychological Review*, 104, pp. 266-300.

- PALMER S. (1977), «Hierarchical Structure in Perceptual Representation», in *Cognitive Psychology*, 9, pp. 441-474.

- PUTNAM H. (1975), *Mind, Language and Reality. Philosophical Papers, Volume 2*, Cambridge University Press, Cambridge (trad. it. a cura di Roberto Cordeschi, *Mente, linguaggio e realtà*, Adelphi, Milano, 1987).

- QUILLIAN M. (1968), «Semantic Memory», in M. Minsky (ed.), *Semantic Information Processing*, MIT Press, Cambridge, Mass., pp. 227-270.

- REDDY R. D., ERMAN L. D., FENNEL R. D., NEELY R. B. (1973), «The HEARSAY specch understanding system: an exemple of the recognition processes», in *Proceedings of the Third Joint Conference on Artificial Intelligence*, Stanford, Ca, pp. 175-183.

- REHLING J. A. (1997), *Automating Creative Design in a Visual Domain*, tech. report n. 113, Center for Research on Concepts and Cognition, Indiana University, Bloomington (IN).

- REHLING J. A. (2001), *Letter Spirit (part two): Modeling Creativity in a Visual Domain*, Ph.D. Dissertation, Indiana University, Bloomington (IN).

- REHLING J. A., HOFSTADTER D. R. (1997), «The Parallel Terraced Scan: An Optimization for an Agent-Oriented Architecture», in *Proceedings of the IEEE International Conference on Intelligent Processing Systems 1997*, Beijing, China.

- ROBINSON H. (2004), «Thought Experiments, Ontology, and Concept-dependent Truthmakers», in *The Monist*, 4, pp. 537-553.

- ROSCH E. (1975), «Cognitive representations of semantic categories», in *Journal of Experimental Psychology: General*, 104, pp. 192-233.

- ROSCH E. (1976), «Basic objects in natural categories», in *Cognitive Psychology*, 8, pp. 382-439.

- ROSCH E., LLOYD B. B. (EDS.) (1978), *Cognition and categorization*, Erlbaum, Hillsdale, NJ.

- RUMELHART D. E., MCCLELLAND J. L. *ET AL.* (1986), *Parallel distribuited processing*, MIT Press, Cambridge (Mass.) (trad. it. *PDP, microstruttura dei processi cognitivi*, Il Mulino, Bologna, 1991).

- RUMELHART D. E., ORTONY A. (1977), «The representation of knowledge in memory», in R. C Anderson, R. J. Shapiro, W. E. Montague (eds.), *Schooling and Acquisition of Knowledge*, Erlbaum, Hillsdale, NJ.

- RYLE G. (1949), *The Concept of Mind*, New Univer Edition, University of Chicago Press, Chicago (trad. it a cura di Ferruccio Rossi-Landi, *Lo spirito come comportamento*, Laterza, Roma-Bari, 1982).

- SANDRI G. (2006), «Mutamenti nella nozione di computazione», in *Preprint*, Dip. Filosofia Univ. di Bologna & CLUEB, Bologna, pp. 199-246.

- SCHANK R. C. (1972), «Conceptual dependency: A theory of natural language understanding», in *Cognitive Psychology*, 3, pp. 552-631.

- SCHANK R. C. (1982), *Dynamic memory. A theory of reminding and learning in computers and people*, Cambridge University Press, Cambridge, Mass. (trad. it. a cura di Alessandra Stragapede, *Memoria dinamica. Una teoria della rievocazione e dell'apprendimento nei calcolatori e nelle persone*, Marsilio, Venezia, 1987).

- SCHANK R. C. (1984), *The Cognitive Computer on Language Learning and Artificial Intelligence*, Addison-Wesley, Reading (trad. it. di Gabriele Noferi, *Il computer cognitivo: linguaggio, apprendimento e intelligenza artificiale*, Giunti, Firenze, 1989).

- SCHANK R. C., ABELSON R. P. (1977), *Scripts, Plans, Goals and Understanding*, Erlbaum, Hillsdale.

- SEARLE J. R. (1980), «Mind, Brains and Programs», in *The Behavioral and Brain Sciences*, 3, pp. 417-457 (trad. it. *Menti, cervelli e programmi, un dibattito sull'intelligenza artificiale*, a cura di G. Tonfoni, CLUP-CLUED, Milano, 1984).

- SEARLE J. R. (1983), *Intentionality: An Essay in the Philosophy of Mind*, Cambridge University Press, Cambridge (trad. it. a cura di Daniele Barbieri, *Della intenzionalità. Un saggio di filosofia della conoscenza*, Bompiani, Milano, 1985).

- SIEGELMANN H. T. (1999), *Neural networks and analog computation: beyond the Turing limit*, Birkhauser, Boston.

- SIMON H. A. (1955), «A Behavioral Model of Rational Choice», *Quarterly Journal of Economics*, 69, pp. 99-18.

- SIMON H. A. (1981), *1980 Procter Lecture: Studying Human Intelligence by Creating Artificial Intelligence*, in «American Scientist», 69, pp. 300-309.

- SIMON H. A. (1987), «Bounded Rationality», in J. Eatwell, M. Millgate, P. Newmann, *The new Palgrave: A Dictionary of Economics*, Macmillan, London and Basingstokes.

- SIMON H. A., KOTOVSKY K. (1963), «Human Acquisition of Concepts for Sequential Patterns», in *Psychological Review*, 70, 6, 1963, pp. 534-546.

- SMOLENKY P. (1988), «On the Proper Treatment of Connectionism», in *Behavioral and Brain Sciences*, 11, pp. 1-77 (trad. it. a cura di Marcello Frixione, *Il connessionismo tra simboli e neuroni*, Marietti, Genova, 1992).

- SUNDMAN J. (2003), «Artificial stupidity», http://www.salon.com/tech/feature/2003/02/26/loebner_part_one/

- SUSSMAN G. J. (1975), *A computer model of skill acquisition*, MIT Press, Cambridge, Mass.

- TREISMAN A., GELADE G. (1980), «A feature-integration theory of attention», in *Cognitive Psychology*, 12, pp. 97-136.

- TULVING E. (1972), «Episodic and semantic memory», in E. Tulving, M Donaldson (eds.), *Organization of memory*, Academic Press, New York, pp. 381-403.

- TURING A. M. (1948), «Intelligent Machinery», rapporto interno del National Physics Laboratory, ora in *Collected Works of A. M. Turing: Mechanical Intelligence*, North Holland, Amsterdam, 1992, pp. 1-27 (trad. it. a cura di Gabriele Lolli, «Macchine intelligenti», in , *Intelligenza meccanica*, Bollati Boringhieri, Torino, 1994, pp. 88-120).

- TURING A. M. (1950), «Computing Machinery and Intelligence», in *Mind*, 59, pp. 433-460 (trad. it. a cura di Nino Dazzi, «Macchine calcolatrici e intelligenza», in V. Somenzi, R Cordeschi (a cura di), *La filosofia degli automi*, Bollati Boringhieri, Torino, 1994, pp. 167-193).

- TURING A. M. (1992), *Collected Works of A. M. Turing: Mechanical Intelligence*, North-Holland, Amsterdam (trad. it. a cura di G. Lolli, *Intelligenza meccanica*, Bollati Boringhieri, Torino, 1994).

- VON NEUMANN J., MORGENSTERN O. (1944), *Theory of Games and Economic Behavior*, Princeton University Press, Princeton.

- WALTZ D. L. (1972), «Generating semantic descriptions from drawing of scenes with shadows», in P. H. Winston (ed.) (1975), pp.19-92.

- WEIZENBAUM J. (1965), «ELIZA, a computer program for the study of natural language communication between man and machine», in *Communication of the Association for Computing Machinery*, 9, pp. 36-45.

- WEIZENBAUM J. (1978), *Computer Power and Human Reason: from Judgment to Calculation*, Freeman, San Francisco (trad. it. *Il potere del computer e la ragione umana:i limiti dell'intelligenza artificiale*, Gruppo Abele, Torino, 1987).

- WINOGRAD, T. (1972), *Understanding Natural Language*, Academic Press, New York.

- WINOGRAD T. (1973), «A procedural model of language understanding», in R. Schank, K. Colby (eds.), *Computer Models of Thought and language*, Freeman, San Francisco.

- WINSTON P. H. (ED.) (1975a), *The psicology of computer vision*, McGraw-Hill, New York.

- WINSTON P. H. (1975b), «Learning Structural Descriptions from Examples», in *idem* (1975a), pp. 157-209.

- WINSTON P. H. (1982), «Learning new principles from precedents and exercises», in *Artificial Intelligence*, 19, pp. 321-350.

- WINSTON P. H. (1986), «Learning by augmenting rules and accumulating censors», in R. S. Michalski, J. G. Carbonell, T. M. Mitchell (eds.), *Machine Learning: An Artificial Intelligence Approach*, Morgan Kaufmann, Los Altos, CA, pp. 45-61.