

9. L'emergenza secondo Douglas Hofstadter: metafore e modelli

9.1 *Menti, cervelli e formiche*

Douglas R. Hofstadter è sicuramente l'emergentista più importante della terza cultura. I risultati delle sue ricerche personali e di quelle del gruppo da egli stesso fondato, il FARG (Gruppo di Ricerca sulle Analogie Fluide), stanno contribuendo all'affermazione del paradigma emergentista in filosofia della mente in due modi principali.

Da un lato, Hofstadter ha reso celebre *la metafora della mente come società* - l'unica metafora che sembra in grado di contrapporsi, per efficacia e pertinenza, a quella della mente come software. D'altra parte, il FARG ha creato una serie di programmi per computer che simulano varie attività mentali: tutti questi programmi sono caratterizzati da un'*architettura emergente* e costituiscono il più grande successo dell'emergentismo.

Non è facile sopravvalutare l'influenza di Hofstadter sugli altri emergentisti della terza cultura, perché nei punti cruciali delle argomentazioni di tali autori ritroviamo proprio le sue idee¹. Tuttavia, per quanto riguarda la filosofia

¹L'influenza maggiore è quella che Hofstadter ha esercitato sui connessionisti. I membri del gruppo PDP hanno ben presenti i suoi lavori, sia quando usano il concetto di emergenza [Rumelhart *et al.* 1986b, 270], sia quando parlano di simboli e subsimboli [Rumelhart - McClelland 1986a, 183]. Smolensky riconosce che l'approccio di Hofstadter non è solo molto vicino alla filosofia connessionista, «*ma è una delle sue fonti principali*» [Smolensky 1988a, 245]; Andy Clark, senza mezzi termini, afferma: «*La mia posizione [...] è in sintonia con quella di Hofstadter (e deve molto ad essa)*» [Clark 1989, 28]. Molto interessante è il fatto che, insieme a Hofstadter e ai connessionisti, anche Daniel Dennett viene talvolta considerato tra gli emergentisti più autorevoli [Varela *et al.* 1991, 29]; però Dennett, nel momento in cui deve rendere concreta la "propria" teoria sull'architettura funzionale della coscienza, fa a sua volta riferimento a uno dei modelli di Hofstadter: «*L'architettura Jumbo di Douglas Hofstadter [...] ha le caratteristiche giuste*» [Dennett 1991, 269]. Infine, non bisogna dimenticare che Hofstadter ha contribuito a lanciare la "sfida della complessità" [Bocchi - Ceruti 1985] e che molti studiosi della complessità si rifanno al suo pensiero (von Foerster lo considera addirittura uno dei principali riferimenti per quanto riguarda la nozione di "autoperatore" [von Foerster 1985, 138]).

dell'emergenza, Hofstadter è a sua volta in debito verso un altro autore: il sociobiologo Edward O. Wilson.

Questo debito intellettuale è molto importante, perché avvicina profondamente l'emergentismo di Hofstadter al *British Emergentism*, vicinanza che dimostra indirettamente la tesi secondo cui *il primo emergentismo e l'emergentismo della terza cultura sono solo fasi diverse del medesimo paradigma filosofico*.

Da Wilson, Hofstadter ha palesemente ereditato alcune idee relative ai comportamenti di massa; inoltre, seguendo Wilson, ha adottato sia il concetto di emergenza, sia la metafora della mente come società di formiche. In questo paragrafo, dopo un breve commento sul concetto di emergenza utilizzato da Wilson, esamineremo brevemente sia la teoria dei comportamenti di massa, sia la metafora societaria.

Come si ricorderà, l'opera principale del primo emergentismo, cioè *The Mind and its Place in Nature* di Broad, fu pubblicata nel 1925, subito prima che esplodesse la "rivoluzione quantistica", la quale ben presto attirò tutta l'attenzione di scienziati e filosofi, a scapito dell'emergentismo. Tuttavia, nel 1926, il concetto di emergenza era ancora assai dibattuto e fu indiscusso protagonista del VI° Congresso Internazionale di Filosofia. Nell'ambito del Congresso numerosi autori intervennero presentando le proprie idee sull'applicabilità del concetto di emergenza [Stephan 1992, 26] e uno di questi interventi fu "Emergent Evolution and the Social" di W. M. Wheeler. Nel suo intervento, Wheeler propose di considerare le «*società animali e vegetali come insiemi emergenti*»², cioè come fenomeni "emergenti" nel senso tecnico-filosofico del primo emergentismo.

Ebbene, Wilson utilizza il concetto di emergenza esattamente nel senso di Wheeler e quindi, per quanto abbiamo appena detto, nel senso del *British Emergentism*. Infatti, dopo aver mostrato la semplicità del comportamento dei singoli insetti sociali (formiche, api, vespe sociali, ecc.), egli afferma:

Il singolo insetto sociale, a confronto col singolo insetto solitario, dispiega schemi comportamentali né eccezionalmente ingegnosi né eccezionalmente complessi. Le considerevoli qualità della vita sociale sono fenomeni di massa che emergono dall'intreccio di

²Wheeler W. M. [1926], "Emergent Evolution and the Social", in Brightman E. S. (a cura di), *Proceedings of the Sixth International Congress of Philosophy*, New York 1926, pp.33-46; citato in Wilson [1971, 420].

questi semplici schemi individuali grazie alla comunicazione. In questo principio risiede la massima sfida e opportunità della sociologia degli insetti. Wheeler, nel suo opuscolo *Emergent Evolution and the Social* [...], anticipò questa proprietà [Wilson 1971, 420].

E' verosimile che il concetto di emergenza, di per sé, non avrebbe riscosso l'attenzione di Hofstadter, di formazione matematico, se Wilson non lo avesse utilizzato nell'ambito della propria *teoria stocastica del comportamento di massa*, che fornisce precise indicazioni per realizzare al computer la simulazione di società "intelligenti" composte da individui non intelligenti: è quest'aspetto che probabilmente ha affascinato Hofstadter. Ma procediamo con ordine.

Nel suo *The Insect Societies* (1971), Wilson spiega che le formiche possono apprendere la strada per uscire da un labirinto con sei vicoli ciechi, in un tempo solo due o tre volte superiore a quello necessario ai ratti [Wilson 1971, 405]; ma, differendo in ciò dai ratti, per una formica questa capacità è anche la massima prestazione individuale possibile. L'apprendimento della formica è ristretto a condizioni speciali e possiede un immediato valore adattivo [Wilson 1971, 406]; ciò che la formica ha appreso non è peraltro utilizzabile per favorire l'apprendimento di nuove situazioni [Wilson 1971, 407]. Le formiche, inoltre, comunicano tra loro mediante pochi e relativamente semplici segnali chimici (ferormoni): «*l'intera organizzazione sociale delle formiche venefiche potrebbe essere mediata da non più di dieci ferormoni*» [Wilson 1971, 410]³.

Confrontata con l'imbarazzante semplicità delle singole formiche, la complessità della struttura a caste e delle "azioni di massa" di una colonia è sbalorditiva. Secondo Wilson, tali fenomeni *emergono* come conseguenza delle interazioni tra le formiche. A questo proposito, la più

importante regola relativa all'azione di massa è che essa normalmente deriva da azioni conflittuali di molte operaie. Le singole operaie prestano soltanto un'attenzione limitata al comportamento delle vicine compagne di nido e sono nettamente inconsapevoli della condizione istantanea della colonia come complesso. Chiunque abbia osservato una colonia di formiche migrante da un sito a un altro di nidificazione ha visto questo principio illustrato in

³Queste drammatiche restrizioni si accordano con i limiti anatomici: non solo il sistema nervoso delle formiche contiene una minuscola percentuale del numero di neuroni che si trovano nel cervello di un vertebrato, ma le ramificazioni dendritiche sono più corte e la loro area recettiva è in proporzione molto minore; la quantità di informazioni trasmissibili è modesta e, a causa della sottigliezza della guaina mielinica, la velocità a cui viaggiano gli impulsi è pari a solo un decimo di quella dei neuroni dei vertebrati [Wilson 1971, 411-2].

modo vigoroso. Mentre alcune operaie sciamano all'esterno trasportando uova, larve e pupe nelle mandibole, altre si affannano a riportarle al nido, e altre ancora corrono avanti e indietro senza alcun carico [Wilson 1971, 421].

Del tutto simili sono le paradossali modalità con cui viene edificato il formicaio:

Operaie di *Formica* e altri tipi di formiche lavorano costantemente a obiettivi opposti nello scavare nidi e nel trasportare materiali per il nido [...]. Anche se queste variazioni antagonistiche appaiono caotiche all'osservazione ravvicinata, il loro risultato finale è quasi invariabilmente un nido di buona fattura che si conforma strettamente al progetto esibito da tutta la specie [Wilson 1971, 421-2].

Si potrebbe ipotizzare che durante l'edificazione del nido le operaie ubbidiscano a precise direttive, *specie-specifiche*, impartite in qualche modo sconosciuto: ma abbiamo la *certezza* che non sia così, perché se in un nido vengono fatte nascere operaie di un'altra specie (dette "invitati"), esse «*vengono accettate dai loro ospiti, ma intraprendono la ricostruzione del nido ospite secondo il piano caratteristico della propria specie. Il risultato della competizione tra ospiti e invitati è un nido ibrido di forma intermedia*» [Wilson 1971, 415].

Un esempio particolarmente impressionante di ordine statistico che si forma da elementi competitivi è l'assetto di marcia delle formiche combattenti. All'interno dello sciame di *Eciton burchelli*, per esempio, regna il disordine: quando le formiche si incontrano, si verificano tutti i tipi di contatti, dallo sfregamento delle antenne alla collisione violenta; dopo il contatto, se retrocedono, spesso si tamponano; oppure cambiano direzione; talvolta accelerano bruscamente, ma possono anche scavalcarsi. Eppure, lo sciame nel suo insieme ha una forma invariante: una ellisse con un fronte (di formiche esploratrici) di 10-15 metri e una profondità di 1-2 metri, che si sposta esattamente alla velocità media di 30 centimetri al minuto [Wilson 1971, 422-3]. Come può emergere una forma così precisa? Chi la decide? Chi emana gli ordini necessari?

La teoria stocastica del comportamento di massa di Wilson è un tentativo di risposta: lo sciame, così come il "progetto" del nido, è un *fenomeno emergente di origine statistica*; ciascuna formica, infatti, reagisce agli stimoli con certi comportamenti, la *probabilità* dei quali è predeterminata geneticamente:

1) il singolo insetto sociale, essendo all'oscuro di buona parte dei fenomeni che si verificano nella colonia cui appartiene, risponde ad hoc agli stimoli in cui via via si imbatte; 2) le risposte e le probabilità del loro verificarsi sono programmate geneticamente in modo che il comportamento di massa della colonia sia efficace nei riguardi delle particolari condizioni ambientali sperimentate dalla specie lungo il tempo evolutivo; 3) il programma si evolve insieme ai mutamenti ambientali, sempre in direzione dell'aumento di efficienza della colonia; 4) i rapporti di distribuzione tra le caste, la struttura in base all'età degli individui nella colonia e anche la comunicazione si evolvono tutti in modo da fornire le risposte e la relativa struttura di probabilità con efficienza maggiore a livello coloniale [Wilson 1971, 426-7].

Secondo la teoria stocastica potremmo associare alle formiche delle matrici di valori che forniscono, in base all'età, alla casta di appartenenza e all'atto precedente, *le probabilità di determinate reazioni a particolari stimoli* (per esempio, una matrice potrebbe fornire la probabilità che una data secrezione larvale venga assorbita da un'operaia di una certa età e con un'immediata esperienza passata) [Wilson 1971, 427]. In accordo con la teoria, quando una formica incontra un oggetto o un'altra formica, si verifica un'interazione stocastica il cui risultato è descritto dalle matrici di probabilità. L'insieme di tutte le interazioni stocastiche, tra le formiche e tra esse e l'ambiente, dovrebbe produrre automaticamente i fenomeni emergenti di cui si è detto.

Il capitolo 16 di *The Insect Societies* si chiude con l'indicazione della realizzazione della *simulazione di una colonia di insetti* come obiettivo da raggiungere negli anni a venire; tale realizzazione utilizzerebbe lo stesso linguaggio e le stesse tecniche matematiche necessarie per realizzare simulazioni di altri sistemi a cui possono essere associati fenomeni emergenti. Uno di tali sistemi, secondo Wilson, potrebbe essere il *cervello*.

Quest'associazione - tra cervello e formiche - non è un'idea di Wilson: egli stesso la attribuisce a Meyer, il quale, già nel 1966, aveva «*posto l'accento su somiglianze tra le coazioni dei neuroni nel cervello e delle operaie in una colonia di insetti*» [Wilson 1971, 595].

Hofstadter ha reso famosa la scoperta di queste somiglianze dedicando uno dei dialoghi più belli del suo *Gödel, Escher, Bach*, il "Preludio e... mirmecofuga", nonché vari passi delle sue opere successive, alla seguente analogia: *la mente sta alla vita sociale del formicaio come il cervello sta all'insieme delle formiche*.

Le “formiche” del cervello sono i neuroni: benché, come ogni cellula vivente, essi siano complessi (confrontati con sistemi semplici o di media complessità), è chiaro che sono *relativamente* semplici (confrontati con il cervello nel suo insieme). Infatti, quando riceve un segnale,

un singolo neurone può rispondere solo in un modo molto primitivo: o scarica o non scarica. E' una quantità d'informazione molto piccola, e di conseguenza per poter convogliare o elaborare grandi quantità d'informazione devono certamente intervenire molti neuroni. Quindi si potrebbe congetturare l'esistenza di strutture più vaste, composte di molti neuroni, che trattano i concetti a livello superiore. Senza dubbio ciò è vero, ma l'ipotesi più ingenua, cioè che per ogni concetto distinto esista un gruppo fisso di neuroni, è quasi certamente falsa [Hofstadter 1979, 368-9].

Analogamente, le singole formiche non possiedono le informazioni che invece possiedono *i gruppi di formiche*; d'altra parte, l'ipotesi secondo cui per ogni informazione complessa (riguardante per esempio il progetto del formicaio) esiste un gruppo fisso di formiche, è a sua volta quasi certamente falsa. Secondo Hofstadter, il fatto davvero cruciale è che tra i neuroni esiste una forma di “comunicazione di massa” analoga a quella di una colonia di formiche, definita da Wilson «*come trasporto entro i gruppi di informazioni che un solo individuo non potrebbe trasmettere a un altro*» [Wilson 1971, 424; Hofstadter 1979, 380].

La metafora della mente come colonia di formiche ci invita a immaginare gruppi di formiche cooperanti e informazioni che passano da un gruppo all'altro; ai livelli di organizzazione superiori, si devono immaginare iper-ipergruppi che trasportano e si passano informazioni che nessun ipergruppo, per non parlare dei gruppi o delle singole formiche, potrebbe mai sognarsi («*image hyperhyperteams carrying and passing information that no hyperteam, not to mention team or solitary ant, ever dreamt of*» [Hofstadter 1982b, 646]).

Le “realizzazioni fisiche” delle informazioni complesse che corrispondono ai concetti sono, nella colonia, gli ipergruppi di formiche e, nel cervello, le reti neuronali. Si tratta, per usare la terminologia di Hofstadter, di *simboli* [Hofstadter 1979, 378]. Tuttavia, a differenza dei simboli postulati dalla teoria computazionale della mente, che sono simboli *passivi*, gli ipergruppi di formiche e le reti neuronali sono simboli *attivi*. Qual è la differenza?

Secondo Hofstadter, il cognitivismo classico, ipotizzando la manipolazione simbolica dei simboli mentali, apre la porta al problema del “manipolatore”: *chi o cosa manipola i simboli?* L’unica risposta sensata è: il sistema nel suo insieme, cioè il cervello. Ma in realtà il cervello non “manipola” nulla; il cervello è - nelle parole di Hofstadter - il mezzo nel quale i simboli fluttuano e nel quale si attivano l’un l’altro. Non c’è alcun manipolatore. C’è semplicemente una vasta collezione di “squadre” - configurazioni di scariche neuroniche che, come squadre di formiche, attivano altre configurazioni di scariche. In altre parole, non solo noi non siamo manipolatori di simboli passivi; piuttosto, è vero che noi *siamo manipolati dai nostri simboli attivi* [Hofstadter 1982b, 648].

Come metafora della mente, la società di insetti è ottima. Ma per gli emergentisti essa è anche più di una metafora: sia la mente, sia la società di insetti (in quanto tale) sono *fenomeni emergenti*. Consideriamo infatti una colonia associata al sistema fisico costituito dall’insieme delle formiche. Le proprietà delle formiche, considerate come sistema *sociale*, sono *nuove* rispetto a quelle delle singole formiche (si pensi ad esempio alla suddivisione in caste) e descrivibili nell’ambito di modelli (quelli della mirmecosociologia) che sono *qualitativamente diversi* da quelli usati per i singoli insetti. Inoltre, secondo la teoria stocastica, l’esatto comportamento delle formiche è *imprevedibile* (anche se la *probabilità* del loro comportamento è innata e quindi *teoricamente* prevedibile). Infine, nelle “azioni di massa” di una colonia non vi è assolutamente *nulla* che dipende dalle singole formiche, le quali possono morire o essere rimosse senza che ciò influenzi minimamente il resto della colonia. In breve, almeno secondo la nostra definizione di emergenza, la *società* di formiche (come tale) è un *fenomeno emergente*.

E’ dunque evidente che qualsiasi altra metafora, da un punto di vista emergentista, non sarebbe altrettanto efficace. Il software, per esempio, è uno dei livelli di descrizione del funzionamento dei *sistemi di media complessità*, ai quali possiamo certamente associare fenomeni innocentemente emergenti, *ma non è esso stesso un fenomeno emergente* (tra le altre cose, non è imprevedibile). La società e la mente, al contrario, sono associate a *sistemi complessi* e, secondo gli emergentisti, sono entrambe fenomeni emergenti: ciò le rende particolarmente adatte, come metafore, l’una all’altra.

9.2 Percezione di alto livello, aloni e slittamenti concettuali

Hofstadter e gli altri ricercatori del FARG hanno dedicato gli ultimi dieciododici anni a scrivere programmi per computer che simulano specifiche attività mentali. Tali programmi hanno anche dei precursori: Seek-Whence e Jumbo. Seek-Whence è un modello di come gli esseri umani scoprono la regola implicita di una successione, mentre Jumbo simula l'elaborazione umana degli anagrammi.

I programmi più recenti del FARG sono: Numbo, una versione numerica di Jumbo; Copycat e Tabletop, che simulano l'elaborazione di analogie in due diversi microdominî (le lettere dell'alfabeto e una tavola virtuale apparecchiata); attualmente il FARG sta lavorando a Letter Spirit, un modello della creatività umana nel microdominio delle lettere dell'alfabeto. Tutti questi programmi hanno una serie di caratteristiche in comune:

Prestazioni minime. Se confrontati con i sistemi esperti dell'IA, i programmi del FARG sono come “capriole di bambini”:

mentre sembra che molti gruppi di ricerca affrontino domini tanto complessi da intimorire le persone più esperte (termodinamica, terrorismo internazionale, fisica atomica, configurazioni di sistemi informatici, progettazione di microcircuiti VLSI, previsioni economiche, e chi più ne ha più ne metta), noi, nel nostro gruppo, trattiamo domini tanto microscopici da far apparire quasi banali i risultati dei nostri programmi. Chi porge la minima attenzione a un bambino che si appresti a tentare, per la prima volta, una capriola accanto a un acrobata provetto che esegue spettacolari volteggi sul trapezio? I nostri progetti, almeno a uno sguardo superficiale, sono l'equivalente del bambino che fa la capriola [Hofstadter 1995c, 184].

Ispirazione funzionalista. I programmi del FARG non sono sistemi esperti, ma non sono neppure modelli “ispirati neuralmente” come le reti dei connessionisti, né modelli “descrittivisti”, cioè modelli di come descriviamo, ad alto livello, le attività cognitive. Sono invece modelli che tentano di *simulare la cognizione riproducendo in forma semplificata, al computer, l'attività subcognitiva*. Si noti la differenza dall'IA connessionista cognitiva: quest'ultima si pone come obiettivo (lontano) quello di *simulare la cognizione riproducendo in forma semplificata, al computer, la struttura e il funzionamento delle reti neurali*. Attualmente, le reti neurali non simulano affatto la cognizione, bensì la percezione di basso livello; tuttavia, il punto è che lo fanno

mediante neuroni logici interconnessi. Nei programmi del FARG non vi è nulla di simile a neuroni e sinapsi: la “base” è *concettuale*, non neurale.

Architettura emergente. Se la mente è un fenomeno emergente, allora i programmi che simulano l’attività mentale dovrebbero avere un’architettura funzionale emergente. Questo è, in sostanza, il *credo* del FARG. Il problema diventa: com’è un’architettura emergente? Come sempre quando si parla di emergenza, anche questa volta si deve immaginare una gamma di possibilità *proprio nel mezzo* della quale si colloca ciò che è relativo all’emergenza. L’architettura di un programma come quelli del FARG si colloca a metà strada tra quella dei programmi dell’IA convenzionale e quella dei modelli stocastici dei connessionisti:

non è simbolica, né connessionistica, né un ibrido tra le due (benché alcuni potrebbero considerarla tale); il programma, piuttosto, ha un tipo nuovo di architettura che si situa tra i due estremi. Essa è *emergente*, nel senso che il comportamento globale del programma emerge come conseguenza statistica di una miriade di piccoli atti computazionali [Hofstadter - Mitchell 1993, 225].

Poiché la descrizione minuziosa dell’architettura emergente è molto complicata, è importante avere in mente un’immagine globale di tale architettura a cui fare costante riferimento per evitare fraintendimenti. Lo “scenario” a cui pensare è quello di uno spazio virtuale in cui tanti piccoli agenti cercano di stabilire dei legami tra gli oggetti del dominio (lettere, numeri, ecc.), ognuno nella propria piccola area di lavoro, senza sapere nulla di tutti gli altri; naturalmente, l’immagine

ricorda un po’ quello che avviene in una colonia di formiche, ed è proprio questo che si vuole suggerire: l’attività globale del formicaio è solo il risultato emergente da tutte le minuscole attività al suo interno [Hofstadter 1995d, 75].

Per ovvie ragioni di spazio, non è purtroppo possibile passare in rassegna tutti i modelli creati dal FARG. Ma, data la loro importanza per l’emergentismo, non possiamo neanche ignorarli limitandoci a qualche considerazione marginale. Probabilmente la cosa migliore da fare è presentare, in breve, il programma più caratteristico e meglio riuscito: Copycat. Prima, però, dobbiamo spendere ancora qualche parola sul *livello dell’attività mentale simulata* dai programmi del FARG e sul

concetto di *concetto*; per chiarire il primo punto, ci sposteremo temporaneamente nel dominio di uno dei precursori di Copycat, cioè Seek-Whence.

Seek-Whence (significa “cerca-da-dove”, ma in inglese si pronuncia come “sequenza”) è un programma che cerca l’origine di una successione finita di numeri interi, così da permettere l’estensione all’infinito dello schema di partenza. Per esempio, consideriamo la successione [Hofstadter 1995d, 43]:

1, 4, 9, 16, 25, 36, 49, ...

Il problema sottoposto a Seek-Whence è stato scoprire “cosa viene dopo” (senza tirare a indovinare!). Ma la domanda che qui ci poniamo è: *che tipo di attività mentale è coinvolta in una simile operazione?* E’ certamente un’attività che va ben oltre la portata del semplice riconoscimento percettivo, perché nessuno può “percepire” direttamente la soluzione del problema (o almeno, nessuno lo può fare per *qualsiasi* successione). L’attività richiesta è quella che il FARG chiama “percezione di alto livello”, cioè un tipo di *elaborazione in parallelo* che, a differenza del riconoscimento percettivo, coinvolge *concetti* [Chalmers *et al.* 1992].

La percezione di alto livello è prettamente umana e non è l’attività mentale simulata dai programmi dell’IA convenzionale. Questi ultimi risolverebbero il problema attraverso una sequenza di *operazioni che non corrispondono a ciò che accade nella mente di un essere umano* impegnato nel medesimo compito; applicherebbero, per esempio, un operatore come “differenze prime” ottenendo, dalla successione di partenza, la successione-figlia « 3, 5, 7, 9, 11, 13, ...» e, da questa, la successione-nipote «2, 2, 2, 2, 2, ...». La regola della successione-nipote è facilmente riconoscibile e, una volta scoperta la regola di una qualsiasi successione derivata, i programmi convenzionali sono in genere capaci di risalire l’“albero genealogico” fino alla regola della successione-madre (nell’esempio, l’n-imo termine è n^2). Nulla di simile avviene nella nostra mente.

Hofstadter ci offre una splendida (anche se lunghissima) descrizione dell’attività percettiva di alto livello che si svolge nella nostra mente quando cerchiamo di scoprire la regola implicita di una successione; dato che è veramente illuminante, vale la pena di riportarla per intero (o quasi).

La successione considerata è:

1, 4, 27, 256, 3125, 46656, 823543, 8388608, ...

All'inizio, i tre numeri sulla destra, così grandi, potrebbero sembrare tanto minacciosi da essere quasi ignorati; l'unica considerazione che potrebbero suggerire è che entrano in gioco operazioni che portano a numeri grandi. Quindi ci si limiterà a studiare ciò che succede nel frammento iniziale: 1, 4, 27, 256, 3125 [Hofstadter 1995d, 55].

1 4 27 256 3125 ...

I primi due termini non dovrebbero attrarre molto l'attenzione, dato che gli interi 1 e 4 non hanno alcuna caratteristica rilevante; lo stesso si può dire del quarto e del quinto termine, dato che gli interi 256 e 3125 sono poco familiari e poco usati [...]. Il 27, invece, avrebbe ottime probabilità di risaltare e occupare la scena, dato che ha una e una sola proprietà importante: è il cubo di 3 e questa definizione gli rimarrebbe «inchiodata». Nulla garantisce, certo, che questa sia una proprietà rilevante; potrebbe anche trattarsi di una pista falsa, ma almeno è una traccia, qualche cosa che ci permette di cominciare la partita. Un effetto secondario di questo primo suggerimento locale potrebbe essere quello di farne scattare uno tematico, di scala più ampia: *descrivere i numeri come potenze*; forse i cubi o forse no, ma comunque potenze. Fra l'altro, ciò combacia con l'intuizione precedente di avere a che fare con operazioni, per così dire, di alta potenza [Hofstadter 1995d, 55-6].

1 4 3³ 256 3125 ...

L'idea delle potenze sembra essere calzante, dati i termini della successione, anche se alcuni aspetti potrebbero essere, all'inizio, poco chiari. Per esempio: come descrivere il numero 1, dato che tutte le potenze di 1 sono 1? Come descrivere 256? Se si considera questo numero nel contesto delle «potenze», affiorano ricordi vaghi a proposito del fatto che esso è una certa potenza, grande, di 2, forse il quadrato di 16. (Questo genere di cose è spesso piuttosto vago, nella mente, anche se i fatti aritmetici sono di certo oggettivi e si possono ricavare dal calcolo, quando occorra). E' molto probabile che sarebbe utile pensare, almeno in via temporanea, a 256 come «alla settima o ottava potenza di 2». E 3125? Dato che finisce con 25, *parrebbe poter* essere una potenza di 5, ma quale? Per fortuna almeno una cosa è certa: 4 si può descrivere come il quadrato di 2 [Hofstadter 1995d, 56].

1 *qualunque cosa* 2² 3³ 2 *circa 7 o 8* 5 *qualcosa* ...

Dato questo insieme di descrizioni temporanee, cominciano ad affiorare alcune tendenze: frammenti di strutture sottintesi dalle descrizioni stesse. La cosa più significativa è il fatto, molesto, che si ha una potenza di 1, una di 2, una di 3 e poi, invertendo la tendenza, ancora una potenza di 2. Per fortuna chiunque abbia un'infarinatura di matematica sa che metà delle potenze di 2 sono anche potenze di 4 [Hofstadter 1995d, 56-7].

1 *qualunque cosa* 2² 3³ 4 *qualcosa* 5 *qualcosa* ...

Così arriva un ordine: «Verificare che 256 sia una potenza di 4». Un calcolo veloce rivela che davvero lo è: in particolare, ne è la *quarta* potenza. Allora si mette questa descrizione al suo posto, scartando la precedente [Hofstadter 1995d, 57].

1 *qualsunque cosa* 2² 3³ 4⁴ 5 *qualcosa* ...

A questo punto balza agli occhi che le descrizioni portano in un'altra direzione, quella della *identità tra esponenti e relative basi*, almeno in tre casi successivi: 2², 3³, 4⁴. Ciò lancia un messaggio, forte e chiaro, a 1: dovrebbe descriversi come 1¹, e non come 1⁰ o 1² o qualsiasi altra potenza di se stesso [Hofstadter 1995d, 57].

1¹ 2² 3³ 4⁴ 5 *qualcosa* ...

Per la medesima ragione (identità tra base ed esponente) sorge la domanda relativa a 3125: «E' per caso definito da 5⁵?». Di nuovo, un rapido calcolo rivela che il sospetto era giusto, e ormai il gioco è fatto [Hofstadter 1995d, 57].

1¹ 2² 3³ 4⁴ 5⁵ ...

Questa lunga descrizione esemplifica il tipo di elaborazione riprodotta dai programmi del FARG. A differenza dell'elaborazione distribuita in parallelo delle reti neurali, essa coinvolge numerosi concetti (nell'esempio: *numero, base, esponente, identità, ecc.*).

E' molto importante avere presente che, quando parlano di *concetti*, i "FARGonauti" immaginano in realtà degli *aloni concettuali*. Nelle parole di Hofstadter:

tutti i concetti nella mente sono circondati da un alone di concetti confinanti. [...] Vicino al nucleo dell'alone concettuale «contattare» ci sono concetti come «telefonare», «andare a trovare», «scrivere». Più in là potrebbero esserci concetti del tipo «sognare di», «comunicare psichicamente» e così via [Hofstadter 1988, 159].

L'esistenza degli aloni concettuali è concretamente testimoniata dalle "scivolate d'alone" che si manifestano attraverso i *lapsus linguae*, come ad esempio la commistione di parole: «Non c'era nemmeno un posto su quel voreo» (in cui *volo* e *aereo* si sono mescolati), «Non urdare!» (mescolanza di *urlare* e *gridare*), ecc. Più strani sono gli errori di sostituzione totale, come questo: «La porta del bagno non si chiude, il rubinetto non funziona bene», dove *rubinetto* ha superato *maniglia* nella gara per essere espressi; entrambi «sono abitanti caratteristici di quello spazio

semantico che comprende impugnature o comandi che ruotano, ma nel contesto rappresentato dal bagno il rubinetto diventa molto più importante» [Hofstadter 1995a, 219].

La sostituzione totale di un concetto con un altro è il fenomeno grazie a cui gli esseri umani possono fare *analogie*. Quando avviene durante l'elaborazione di un'analogia, chiameremo tale fenomeno "slittamento concettuale", intendendo con ciò

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 1995a, 216].

9.3 L'architettura emergente di Copycat

L'abilità che hanno i concetti di slittare uno nell'altro è responsabile della *fluidità mentale*: la proprietà fondamentale del pensiero secondo il FARG⁴.

Il programma Copycat (significa "copione", nel senso di imitatore pedissequo) vuole essere, più che un "semplice" simulatore di analogie, un vero e proprio modello della *fluidità mentale*. Tuttavia, esso è stato effettivamente concepito per trattare le analogie in un dominio limitato. Esempi tipici di problemi da sottoporre a Copycat sono i seguenti [Hofstadter - Mitchell 1993, 225-6]:

- 1) La stringa di lettere **abc** viene modificata in **abd**; cosa diventa **ijk** se la si modifica "allo stesso modo"?
- 2) La stringa di lettere **aabc** viene modificata in **aabd**; cosa diventa **ijkk** se la si modifica "allo stesso modo"?

⁴La parola «fluidità» non è accidentale: vuole evocare l'immagine dell'acqua. Ora, com'è noto, ogni goccia d'acqua è costituita da migliaia di miliardi di molecole di H₂O che interagiscono stocasticamente e, tramite il debole legame a idrogeno, formano piccoli "aggregati sfarfallanti" instabili e mutevoli. Ebbene, la fluidità e le altre proprietà dell'acqua sono dovute proprio al comportamento di tali aggregati sfarfallanti. *Settant'anni dopo la scelta di Broad di esemplificare le proprietà emergenti tramite le proprietà dell'acqua [Broad 1925, 62-3], e nonostante l'evoluzione del loro pensiero, per gli emergentisti quella dell'acqua è ancora «l'immagine più appropriata per dare un'idea della nostra filosofia secondo cui le caratteristiche di fluidità del pensiero, familiari e in apparenza stabili, emergono come conseguenza statistica di una miriade di piccole azioni indipendenti e subcognitive che si svolgono in parallelo. I concetti possiedono questa fluidità e le analogie ne sono una manifestazione essenziale» [Hofstadter 1995b, 18].*

La maggioranza delle persone risolverebbe il primo problema rispondendo *ijl*, interpretando l'operazione $abc \rightarrow abd$ come la *sostituzione della lettera all'estrema destra con la lettera seguente dell'alfabeto*. Solo una piccolissima percentuale di persone potrebbe essere così pedissequa da rispondere *ijd* (sostituzione della lettera all'estrema destra con la lettera *d*), *ijk* (sostituzione di tutte le *c* con *d*) o addirittura *abd* (sostituzione della struttura con *abd*) [Hofstadter - Mitchell 1993, 226].

Il secondo problema è molto più sottile. In questo caso, infatti, l'interpretazione di prima - cioè la *sostituzione della lettera all'estrema destra con la lettera seguente dell'alfabeto* - appare rozza: la soluzione $ijkk \rightarrow ijkl$ è molto pedissequa. Per qualche ragione, considerare insieme le due *k* sembra più "naturale": l'operazione $aabc \rightarrow aabd$ potrebbe essere interpretata come la *sostituzione del gruppo all'estrema destra con il suo successore alfabetico* e l'analogia potrebbe diventare $ijkk \rightarrow ijll$. La nuova interpretazione, meno pedissequa della precedente, è dovuta a uno "slittamento concettuale", dal concetto di *lettera* a quello di *gruppo di lettere*.

Tuttavia, la "corrispondenza di identità" tra *aa* e *kk* - se la si nota - può indurre un secondo slittamento, da *più a destra (c, kk)* a *più a sinistra (aa, i)*, e quindi una nuova soluzione: $ijkk \rightarrow jkkk$. E' però probabile che il secondo slittamento ne causi un terzo, da *successore a predecessore*, e che la soluzione diventi: $ijkk \rightarrow hjkk$. Come mai il secondo problema può provocare tutti questi slittamenti concettuali? La ragione è che l'elaborazione dell'analogia è avvenuta sotto "pressione": il raddoppiamento delle *a* e delle *k*. A proposito di questo raddoppiamento di lettere,

si potrebbe considerarlo un dispositivo per generare «enfasi», in modo che il limite sinistro della prima stringa e quello destro della seconda risaltino e in un certo senso si «attraggano» a vicenda [Hofstadter - Mitchell 1993, 227].

Copycat risolve i due problemi dando tutte le risposte che abbiamo proposto, con frequenze proporzionali a quanto le soluzioni sono "naturali" (secondo una valutazione umana). Ma, cosa molto più importante, Copycat risolve problemi di questo tipo *nello stesso modo* in cui li risolve un essere umano. Per poter ottenere questo risultato, è necessaria l'architettura funzionale che ora descriveremo.

A differenza di una rete neurale, che “occupa” un unico spazio virtuale, Copycat può essere descritto solo facendo riferimento a ben *tre* spazi virtuali: la rete di slittamento, lo spazio di lavoro e l’appendicodici. (E’ subito chiaro, pertanto, che l’architettura di Copycat è molto più complicata di quella delle reti neurali.) Esaminiamo questi spazi.

a) *La rete di slittamento*. Il primo spazio virtuale è simile a una rete neurale preventivamente addestrata: si tratta di una rete concettuale, dove - al posto dei neuroni logici - vi sono *concetti* e - al posto delle connessioni - *legami concettuali*. I concetti di Copycat sono sessanta, tra cui: *a, b, c, ..., x, y, z, lettera, successore, predecessore, primo in ordine alfabetico, ultimo in ordine alfabetico, posizione nell’alfabeto, sinistra, destra, direzione, estremo destro, estremo sinistro, centrale, posizione di stringa, gruppo, gruppo di identità, gruppo di successività, gruppo di precedenza, lunghezza del gruppo, 1, 2, 3, identità, opposto*, ecc. [Hofstadter - Mitchell 1993, 232].

Mentre i processori di una rete neurale hanno un unico valore di scarica (variabile), a ciascun concetto della rete di slittamento vengono associati *due* valori: la *profondità concettuale* (costante) e lo *stato di attivazione* (variabile); lo stato di attivazione di un concetto diminuisce automaticamente nel corso dell’elaborazione (e la velocità di decadimento è inversamente proporzionale alla profondità concettuale).

Così come in una rete neurale troviamo i pesi sulle connessioni, nel programma di Copycat troviamo un valore, chiamato *distanza concettuale*, per ciascun legame concettuale della rete di slittamento. Ma, a differenza dei pesi sulle connessioni, che dopo l’apprendimento non variano più, la distanza concettuale varia proprio durante la simulazione; tuttavia, ogni legame è caratterizzato da un valore caratteristico *costante* a cui la distanza tende sempre a ritornare (è come se la rete fosse “elastica”).

L’attivazione di un concetto attiva tutti gli altri concetti della rete in misura inversamente proporzionale alla loro distanza: questo fatto rende i concetti davvero simili ad *aloni* (o nuvole concettuali) e

suggerisce di pensare alla rete di slittamento non tanto in termini di una rete ben definita di punti e di linee, ma piuttosto come uno spazio in cui molte nubi diffuse si sovrappongono in modo intricato e variabile nel tempo [Hofstadter - Mitchell 1993, 235].

Minore è la distanza tra due concetti, maggiore sarà la probabilità di uno slittamento tra essi.

b) *Lo spazio di lavoro.* Nella rete di slittamento non avviene alcuna elaborazione (diversamente da una rete neurale, che è ovviamente il luogo dove si svolge l'elaborazione). L'elaborazione - cioè la simulazione della percezione di alto livello - avviene in uno spazio virtuale apposito, analogo al *citoplasma* delle cellule: lo spazio di lavoro. All'inizio della simulazione, esso contiene solo i dati grezzi del problema, cioè le lettere delle stringhe di partenza, e nessuna relazione tra esse. Tuttavia, ben presto lo spazio di lavoro si riempie di *strutture percettive*, cioè di strutture virtuali che rappresentano le relazioni tra le lettere (o tra i gruppi di lettere, o tra le stringhe di lettere). Ogni struttura percettiva corrisponde a qualche concetto della rete di slittamento. Chi costruisce tali strutture? Piccoli agenti (processori che funzionano in parallelo), chiamati "codicelli", che vivono nel terzo spazio virtuale di Copycat: l'appendicodici.

Quando Copycat, per esempio, affronta il problema 2, è molto probabile che tra le lettere di *ijkk* i codicelli costruiscano in fretta le seguenti strutture: l'*identità* tra le lettere *k* e la *successività* tra *i* e *j*; le lettere, così unite, vengono percepite come gruppi. Col tempo, vengono costruite strutture tra gruppi di lettere: per esempio, sempre nel problema 2, i codicelli presto si accorgeranno dei due gruppi di identità *aa* e *kk* e costruiranno una struttura percettiva che concretizzi il loro legame: una *corrispondenza di identità*.

Ogni struttura percettiva costruita nello spazio di lavoro invia una carica notevole di attivazione al concetto corrispondente nella rete di slittamento; l'effetto di questa attivazione durerà per un periodo che dipende dal tasso di decadimento del concetto (e a sua volta, il tasso di decadimento dipende dalla profondità concettuale). Poiché nella rete di slittamento l'attivazione di un concetto attiva anche quelli del suo alone, può verificarsi uno slittamento concettuale che influenza l'elaborazione successiva nello spazio di lavoro. Dunque, tra rete di slittamento e spazio di lavoro si verificano continue azioni e retroazioni: *ciò «significa che esiste un ciclo di retroazione tra l'attività percettiva e quella concettuale»* [Hofstadter - Mitchell 1993, 244].

c) *L'appendicodici*. Il terzo spazio virtuale di Copycat è il luogo dove si trovano i codicelli in attesa di partecipare all'elaborazione (cioè alla costruzione di strutture percettive nello spazio di lavoro). *I codicelli non sono altro che le "formiche" di Copycat*. Essi si comportano come le formiche, sono poco intelligenti come le formiche e sono divisi in caste come le formiche.

A ciascun codicello-formica viene associato un *valore di urgenza*, un numero che determina la *probabilità* di partecipare all'elaborazione (nell'istante successivo). La scelta dei codicelli, compiuta dal programma di simulazione, è dunque probabilistica - e ciò rende stocastica l'elaborazione (così com'è stocastico il carattere del comportamento delle formiche). Si noti che i codicelli non sanno nulla l'uno dell'altro, né della condizione istantanea dell'elaborazione nel suo complesso; nessuno di loro è importante: conta solo il loro effetto collettivo.

I codicelli sono divisi in caste: esploratori e fattivi. Gli esploratori, se chiamati a partecipare all'elaborazione, propongono le strutture percettive e creano i fattivi, che le costruiscono (ed esse vanno a riempire lo spazio di lavoro). L'azione degli esploratori, chiamata "scansione parallela a schiera", non può non far pensare alle incredibili marce delle formiche:

Una metafora che può illustrare la scansione parallela a schiera è data dall'immagine di una lunga colonna di formiche che marcia in una foresta guidata da una schiera di esploratori, i quali effettuano piccole incursioni casuali in tutte le direzioni (pur esplorandone alcune con zelo e profondità maggiori) e poi ritornano a riferire: l'effetto collettivo di questi «sondaggi» determinerà la direzione seguita dalla colonna intera; dato che gli esploratori compiono il loro lavoro costantemente, la direzione di marcia della colonna subirà continui, seppur lievi, aggiustamenti [Hofstadter - Mitchell 1993, 248].

La popolazione dell'appendicodici, durante la simulazione, è assai mutevole. Intanto, tutti i codicelli, una volta entrati in azione, vengono eliminati; per non svuotare l'appendicodici, però, il programma di simulazione crea continuamente codicelli esploratori chiamati "osservatori" (o *bottom-up*). Inoltre, cosa più importante, *i concetti attivi della rete di slittamento creano altri codicelli esploratori chiamati "cercatori" (o top-down): è questo il modo in cui la rete concettuale influenza l'elaborazione percettiva*. I codicelli esploratori, a loro volta, creano tutti quelli fattivi.

Come si vede, l'architettura di Copycat è molto complicata. Ma cosa succede effettivamente quando il programma affronta un problema di analogia? La situazione iniziale è la seguente:

- a) Nella rete di slittamento è attivato un insieme fisso di concetti poco profondi;
- b) Nello spazio di lavoro ci sono solo le lettere, senza strutture percettive;
- c) Nell'appendicodici ci sono solo codicelli osservatori in attesa di chiamata.

I primi codicelli chiamati a partecipare all'elaborazione propongono le prime strutture (di solito le *identità*) e creano i fattivi che le costruiscono. Cominciano così ad attivarsi i concetti della rete di slittamento, che a loro volta producono codicelli *top-down*. Questi ultimi rappresentano le “pressioni dall'alto”, le quali dipendono dal contesto specifico del problema. I codicelli *top-down*, insieme a quelli *bottom-up*, esplorano lo spazio di lavoro osservando e cercando relazioni. Nuove strutture possono poi aggiungersi o sostituirsi a quelle precedenti; e così via.

Le “tendenze generali” a cui si assiste nel corso della simulazione sono [Hofstadter - Mitchell 1993, 252]:

- a) Nella rete di slittamento, *i concetti attivati sono sempre più profondi*;
- b) Nello spazio di lavoro, si va *da uno stato senza strutture ad uno con molte strutture*;
- c) Nell'appendicodici, la tendenza è passare *da una popolazione bottom-up ad una top-down*.

E' importante comprendere che il contenuto dello spazio di lavoro rappresenta il *punto di vista* effettivo di Copycat (i singoli codicelli non hanno alcun punto di vista). La scansione parallela a schiera, però, permette di esplorare, in un certo senso, *potenziali* punti di vista alternativi, i quali “fremono probabilisticamente” per imporsi: e se gli esploratori trovano un punto di vista potenziale

abbastanza promettente, creano codicelli fattivi che quando agiranno tenteranno di realizzare, nello spazio di lavoro, questo punto di vista alternativo; si scatenerà allora una «lotta» tra il titolare e il nuovo venuto; l'esito sarà deciso in maniera probabilistica [Hofstadter - Mitchell 1993, 248].

Infine, in Copycat, vi è una variabile che misura il grado di casualità usato dal programma nel “prendere decisioni” (quali codicelli dovranno entrare in azione, quale

di due punti di vista rivali vincerà la contesa, ecc.). Tale variabile è chiamata *temperatura* ed è analoga alla temperatura computazionale dei modelli stocastici dei connessionisti; la differenza è che la temperatura di Copycat *misura* il grado di casualità, mentre quella dei modelli stocastici *controlla* la casualità (influenzando la probabilità che i processori si attivino). La temperatura può essere interpretata come un'indicazione approssimativa dell'auto-valutazione dell'esecuzione: minore è la temperatura, migliore è il giudizio che Copycat dà alla soluzione da lui stesso elaborata [Hofstadter - Mitchell 1993, 251].

Concludiamo riportando, senza commentarli, i risultati di 1000 esecuzioni del programma, tutte relative al problema 2 (il valore della temperatura riportato è la media delle temperature finali delle esecuzioni pervenute allo stesso risultato) [Hofstadter - Mitchell 1993, 260]:

aabc → *aabd* *ijkk* → ?

Soluzione	Frequenza	Temperatura
<i>ijll</i>	612	29
<i>ijkl</i>	198	49
<i>jkkk</i>	121	47
<i>hjkk</i>	47	19
<i>jkkk</i>	9	42
<i>ijkd</i>	6	57
<i>ijdd</i>	3	46
<i>ijkk</i>	3	69
<i>djkk</i>	1	58

Si noti la presenza di varie soluzioni bizzarre (rare e caratterizzate da alta temperatura): Copycat è dotato di un'architettura

che nella maggior parte dei casi gli permette di evitare risposte strane o in apparenza folli, ma che non le evita del tutto; è molto importante che al programma (come alle persone) sia data la *possibilità* di seguire percorsi rischiosi (magari anche folli), affinché abbia la flessibilità per seguire quelli *illuminanti*, ma esso deve anche evitare quelli sbagliati, almeno nel maggior numero dei casi [Hofstadter - Mitchell 1993, 257].

9.4 Osservazioni conclusive

Copycat è un programma davvero affascinante. Non solo simula in maniera molto sofisticata il nostro modo di elaborare analogie, ma la sua flessibilità nell'affrontare i problemi è realmente simile a quella che caratterizza la mente umana. In un certo senso, le sottigliezze dei problemi più difficili hanno costituito per Copycat una specie di test del pensiero (*test di Turing*) “in miniatura”, dal quale si può dire che sia uscito vittorioso [Hofstadter - Mitchell 1993, 270]. Sottoposto per esempio ai problemi seguenti:

- $rst \rightarrow rsu, xyz \rightarrow ?$ (a differenza di t, z non ha alcun successore nell'alfabeto);
- $abc \rightarrow abd, mrrjjj \rightarrow ?$ (a differenza di $abc, mrrjjj$ è formato da gruppi di più lettere);

Copycat ha risposto con audacia e ha anche manifestato certe tendenze non previste nel programma, che rivelano la sua “personalità” (metaforicamente parlando).

Siamo convinti che Copycat sia *veramente* un'efficace riproduzione di alcuni aspetti importanti della fluidità mentale, della quale le analogie sono (forse) la manifestazione essenziale. La notevole complessità dell'architettura funzionale necessaria per questa riproduzione è però scoraggiante, se si pensa al fatto che Copycat (come tutti i programmi del FARG) opera in un dominio estremamente limitato.

La complessità necessaria per simulare analogie *tra dominî diversi* (cosa del tutto naturale per la mente) è sicuramente molto maggiore. Ma *quanto* maggiore? E' probabile che ben pochi ricercatori abbiano un'idea precisa della risposta da dare a questa domanda, perché per molti, troppi anni l'IA è sfuggita al problema della *complessità reale della mente*. Anzi, si può dire che solo oggi, grazie al connessionismo e al FARG, cominciamo a capire *quanto* sia ancora lunga la strada da percorrere.

Due osservazioni vanno qui aggiunte per concludere il nostro discorso. La prima osservazione riprende la distinzione che abbiamo tracciato tra spiegazione oggettiva, comprensione soggettiva e riproduzione esplicativa (§7.1). Da nessuna interpretazione dei programmi del FARG - neppure di Copycat - ci si dovrebbe aspettare una *spiegazione* dei fenomeni simulati, cioè il raggiungimento di una

conoscenza completa e completamente soddisfacente di essi. Hofstadter ha le idee molto chiare a questo proposito:

per quanto apprezzabile, questa ricerca non arriva a dare una spiegazione completa dei fenomeni che studia; e dopo tutto, nessun lavoro scientifico dice l'ultima parola sul tema che tratta - meno che mai nel caso delle scienze cognitive, le quali cominciano appena adesso a svelare la complessità della mente [Hofstadter 1993, 333].

Riteniamo che questa precisazione sia importante, perché influenza il nostro modo di interpretare l'applicazione del concetto di emergenza ai fenomeni mentali. Tale applicazione *non* viene compiuta nell'ambito della ricerca di una spiegazione completa della mente (spiegazione che potrebbe anche essere impossibile per ragioni di principio).

La chiave di lettura con cui interpretare l'applicazione del concetto di emergenza alla mente dovrebbe piuttosto essere quella secondo cui tale applicazione fa parte del tentativo di capire la mente per mezzo di modalità conoscitive diverse dalla spiegazione, come la comprensione e la riproduzione. La metafora della mente come colonia di formiche, da parte sua, contribuisce significativamente alla *comprensione intuitiva* di molti fatti relativi alla mente: per esempio, il fatto che *le parti di un sistema dinamico non devono necessariamente essere intelligenti affinché l'intelligenza si possa associare al sistema nel suo insieme*. Un programma come Copycat, d'altra parte, riproduce altri fatti relativi alla mente che, per il fatto stesso di essere stati riprodotti, possono finalmente essere "capiti": per esempio, il fatto che *alla base della fluidità mentale non solo vi sono, ma probabilmente vi devono essere processi stocastici, cioè processi che, singolarmente considerati, sono casuali*.

La seconda ed ultima osservazione riguarda l'interpretazione neurologica dell'architettura emergente di Copycat. A quale struttura neuronale corrisponde un alone concettuale? A cosa corrisponde lo spazio di lavoro? A cosa i codicelli e a cosa l'appendicodici? La risposta a tutte queste domande è la medesima: *francamente, nessuno lo sa*.

Questa lacuna, ovviamente, è molto grave. Tuttavia, non è un problema *così* grave da dover pensare che la strada intrapresa da Hofstadter e colleghi sia sbagliata. Piuttosto, è proprio il tipo di ricerche di gruppi come il FARG che ci permette di

precisare quali siano i problemi ancora irrisolti e a quali domande dobbiamo ancora rispondere. In altre parole, è molto meglio che un problema aperto sia quello della corrispondenza neuronale dei codicelli che non quello, fuorviante in partenza, della localizzazione cerebrale di un chimerico manipolatore di simboli mentali.

Concordiamo comunque con i membri del FARG: i loro modelli sono importanti perché esplorano

un terreno intermedio della simulazione cognitiva, posto tra quello dei sistemi simbolici di alto livello e quello dei sistemi connessioneistici di basso livello; la ragione per cui si sostiene questo approccio è che il suo livello intermedio è al momento il più utile per trattare la natura fluida dei concetti e delle percezioni, aspetti centrali della mente che emergono con la massima chiarezza nel fare analogie [Mitchell - Hofstadter 1993, 318].

Pertanto, nonostante il problema dell'interpretazione neurologica, riteniamo che la creazione di programmi come Copycat, cioè ad architettura funzionale emergente, sia da guardare come un grande successo scientifico e una significativa affermazione dell'emergentismo in filosofia della mente.

Conclusioni

Una “soluzione” emergentista al problema mente-corpo?

Nell'elaborare questa tesi abbiamo seguito la convinzione secondo cui, per affrontare la tematica della mente come emergenza, è indispensabile confrontarsi con i problemi relativi non tanto ad un unico concetto - il concetto di emergenza - quanto ad un ben definito e articolato *paradigma filosofico*: l'emergentismo.

Nel corso del lavoro si è cercato di mostrare implicitamente quanto fosse importante questa impostazione, sia per la prospettiva storica, sia per quella epistemologica. In questo capitolo conclusivo torneremo sul problema mente-corpo per rispondere alla seguente domanda: l'emergentismo (come tale, cioè indipendentemente dal pensiero di singoli autori) offre una soluzione accettabile a tale problema?

Popper ha ampliato il problema mente-corpo spostando l'attenzione verso il rapporto io-cervello; i sostenitori della teoria computazionale della mente hanno ampliato a loro volta il problema facendo irrompere in esso la questione del rapporto tra mente, cervello, io o sé e il computer (inteso come hardware + software). Dopo questo doppio ampliamento, il problema mente-corpo si presenta come un triplice interrogativo:

1. Qual è la natura della mente?
2. Qual è la natura dell'io (o del sé)?
3. Quali rapporti intercorrono tra mente, io (o sé), cervello e computer?

La conclusione a cui siamo giunti al termine della nostra ricerca è la seguente: *l'emergentismo propone una soluzione al problema della natura della mente, ma non al problema della natura dell'io; per quanto riguarda il problema dei rapporti tra mente, io, cervello e computer, l'emergentismo propone delle idee interessanti per i*

rapporti in cui non è coinvolto l'io. Il problema non può dirsi risolto per quanto riguarda gli altri rapporti.

Le idee che gli emergentisti hanno prodotto relativamente ai rapporti che intercorrono tra mente, cervello e computer sono state illustrate dettagliatamente negli ultimi due capitoli della nostra ricerca e non ci sembra necessario rivederle. Intendiamo invece commentare le altre proposte.

L'osservazione più importante è che la soluzione emergentista al primo problema (la natura della mente) non è separabile da quella proposta come soluzione al problema del rapporto mente-cervello: *la natura della mente sta nel suo rapporto fisico con il cervello*. Infatti, secondo l'emergentismo, la mente è un fenomeno emergente: *cioè un fenomeno indissociabile dal sistema nervoso, avente le caratteristiche di cui alla definizione di emergenza*.

Si noti che questa è una risposta diretta ed esplicita al problema posto. Non dice che cosa la mente non è, non dice qualcosa su ciò che si sa (o non si sa) della mente: ci dice proprio qual è, dal punto di vista emergentistico, la natura della mente. Eppure sulle prime non convince. Abbiamo infatti la tentazione di dire: sì, va bene - è un fenomeno emergente - *ma cioè cos'è?*

La domanda «che cos'è la mente?», iterata nonostante la risposta, assomiglia molto a ciò che il filosofo Peter Bieri definisce “ruota tibetana della preghiera” [Dennett 1991, 314]. La ruota tibetana di Bieri è infatti una qualsiasi richiesta di spiegazione, iterata senza riguardo a quale teoria sia stata effettivamente proposta come spiegazione. Nel caso della mente, è molto probabile che la tentazione di iterare la domanda «che cos'è?», sia legata alla ben nota difficoltà, per l'immaginazione umana, di trattare con i *processi* più che con le *cose*.

La mente umana tende, sempre e comunque, alla *reificazione* (e quella dei filosofi addirittura all'ipostatizzazione!). Pertanto, osservando un fenomeno nel quale è coinvolto un processo, si ha la sensazione di aver capito il fenomeno solo quando si trova *la cosa responsabile del processo*. La tendenza alla reificazione ha comportato, probabilmente, un vantaggio dal punto di vista evolutivo, perché in natura i pericoli (per esempio i predatori) sono più simili a “cose” che non a “processi”; tuttavia, questo vantaggio evolutivo costituisce uno svantaggio per l'immaginazione, almeno quando ci si occupa della natura della mente. Infatti, se si rispondesse alla domanda

«che cos'è la mente?» dicendo: una sostanza mentale, la nostra tendenza alla reificazione sarebbe appagata e avremmo la sensazione di aver *davvero* risposto.

Questa sensazione di appagamento, così almeno sembra, *non si ha* quando rispondiamo: la mente è un fenomeno emergente. Tuttavia, noi dobbiamo superare (per quanto possibile) le tendenze ancestrali della nostra psiche e attenerci all'indagine razionale.

Le possibilità, da un punto di vista puramente razionale, sono solo due: o si accetta l'emergenza come una categoria che definisce una classe di fenomeni *sui generis*, oppure la si rifiuta una volta per tutte. Ma, e questo è il punto, *se la si accetta*, si deve anche accettare che la mente possa essere uno dei fenomeni definiti da tale categoria. E, in questo caso, si deve accogliere l'emergenza della mente come una possibile soluzione al problema della natura della mente. In altri termini: fino a un certo punto è sicuramente legittimo iterare la domanda «che cos'è?», *ma arriva un momento in cui ci si deve fermare*. Probabilmente, se la categoria usata per la mente è troppo ampia (se comprende, cioè, una quantità smisurata di fenomeni), allora non è ancora arrivato il momento di fermarsi. D'altra parte, se la categoria usata è troppo stretta - e in particolare se è un "cambio di etichetta", cioè se comprende un unico fenomeno - allora si è andati troppo oltre e occorre tornare indietro. Qual è il momento giusto per fermarsi? Secondo chi scrive, il momento giusto arriva quando *la categoria usata per la mente contiene la quantità più piccola possibile di fenomeni senza essere un cambio di etichetta*. L'emergenza ha proprio questa estensione.

Se ammettiamo sia l'emergenza come categoria di fenomeni *sui generis*, sia la mente come emergenza, dobbiamo anche accettare la possibilità che la natura della mente non sia sostanziale ma processuale. E' questo il nocciolo della celebre "obiezione della passeggiata" di Hobbes. La mente è un fenomeno *emergente*, fondato sul fatto che il cervello è *pensante* senza che i suoi pensieri siano una cosa - così come qualcuno è *passeggiante* senza che la sua passeggiata sia una sostanza: si tratta di fenomeni di natura processuale. L'anti-intuitività di questi fenomeni è palese per chiunque. Tuttavia, è probabile che molti aspetti di questa anti-intuitività siano legati, più che alla mente, al fenomeno che chiamiamo «io» (o «sé»).

Ora, a questo proposito, quello che bisognerebbe evitare è un errore di questo tipo: rifiutare soluzioni parziali a causa della difficoltà maggiore. Se l'emergentismo

offre una soluzione *accettabile* al problema della natura della mente, allora la si dovrebbe *accettare*. E' assurdo ripartire sempre da capo solo perché non si riesce a risolvere il problema mente-corpo (doppiamente ampliato) tutto in una volta. Se, com'è molto probabile, è semplicemente impossibile trovare una soluzione del tipo "tutto o niente" - allora rifiutare i piccoli passi significa condannare la filosofia della mente a non superare mai il problema mente-corpo.

Quello che proponiamo è: (1) ammettere la proposta emergentista come soluzione al problema della natura della mente benché essa non sia *assolutamente* una soluzione al problema della natura dell'io (o del sé) e (2) trasformare il problema mente-corpo nel seguente interrogativo: *in che modo* la mente, intesa come fenomeno emergente, può sviluppare il senso del sé?

Questa proposta, indirettamente, è una critica rivolta verso le contrapposizioni inutili che costano tante energie ai ricercatori che si occupano del problema mente-corpo; è convinzione di chi scrive, infatti, che molti aspetti del problema siano spesso affrontati con un atteggiamento ormai vetusto, basato su dicotomie superate (dualismo sì/dualismo no, libertà/determinismo, pensare per immagini/pensare per parole, ecc.) e soprattutto su metodi antiquati (l'introspezione, gli esperimenti mentali, le analogie tra sistemi appartenenti a ordini di complessità diversi). Occorre, secondo chi scrive, andare avanti, con un atteggiamento di convergenza costruttiva, senza alcun puntiglio tecnicistico, verso una teoria positiva della mente dalla quale (ri)partire per affrontare i problemi più difficili, come la natura dell'io. Ma su quale base tentare una simile convergenza costruttiva? Ci sembra che l'emergentismo abbia i requisiti migliori: la sua posizione nel panorama filosofico contemporaneo, intermedia tra varie correnti, ne fa un paradigma "naturalmente" privo di quelle tesi estreme che di solito dividono il fronte della ricerca.

Ovviamente la convergenza costruttiva può riguardare solo la "terza via" compresa tra il dualismo interazionista di Eccles e l'eliminativismo *forte* verso la mente: queste due posizioni sono ovviamente incompatibili con l'emergentismo. Inoltre, non è nello "spirito" della maggioranza degli emergentisti assumere atteggiamenti normativi - per cui l'adozione dell'eliminativismo forte verso la psicologia del senso comune, da parte di un autore vicino all'emergentismo come Churchland, è certamente destinata a rimanere un caso isolato. Esclusi questi estremi,

però, l'emergentismo risulta piuttosto "ecumenico" (è per esempio compatibile con l'eliminativismo *debole* verso la mente e verso la psicologia del senso comune).

Un discorso a sé stante va fatto per i filosofi che studiano la relazione *logica* tra mente e corpo. L'emergenza è considerata una relazione fisica, non logica. Ci si può chiedere: dal punto di vista logico-formale, potrebbe la mente, rispetto al sistema nervoso, essere *sopravveniente*? Chi scrive ritiene che l'emergentismo potrebbe rispondere affermativamente: nulla, nel paradigma emergentista, impedisce di considerare il rapporto mente-cervello in modo più formale rispetto a quanto facciano solitamente gli emergentisti. Nell'ambito di una simile formalizzazione potrebbe risultare estremamente pertinente interpretare l'emergenza come una relazione di sopravvenienza. La facile compatibilità del principio di eredità causale enunciato da Kim con l'emergentismo della terza cultura suggerisce addirittura che l'emergenza potrebbe essere formalizzabile come una relazione di sopravvenienza *forte*. Le ricerche sul concetto di sopravvenienza sono tuttora in corso e non sappiamo ancora come evolverà la questione.

Chi scrive crede nella possibilità di trattare l'emergentismo non come l'ennesimo paradigma filosofico che si contrappone a tutti gli altri, bensì come la base per tentare una convergenza reale. Convergenza filosofica, ma non solo: la distanza epistemologica dell'emergentismo dalla biologia è minima e ciò lo rende particolarmente interessante anche agli occhi degli scienziati "duri".

La teoria gerarchica del reale e il problema del dualismo

Esaminando la teoria gerarchica del reale abbiamo visto che il livello mentale (o psicologico) dovrebbe essere considerato un *livello di analisi*. Qui faremo qualche osservazione conclusiva a tale proposito.

L'emergentismo contemporaneo "vede" il mondo come una straordinaria gerarchia di livelli di complessità organizzazionale che evolve nel tempo; ogni livello corrisponde a una classe di *sistemi fisici* caratterizzati da: una certa complessità, specifiche proprietà strutturali e funzionali, un determinato comportamento fisico e lo specifico grado di conoscenza che ne abbiamo; discipline diverse studiano i sistemi ai

diversi livelli di questa gerarchia: la fisica, la chimica, la biologia, l'ecologia, la sociologia e le loro innumerevoli ramificazioni. E' probabile, inoltre, che i sistemi appartenenti ad un livello di organizzazione siano apparsi, nel corso dell'evoluzione, *prima* dei sistemi che appartengono al livello superiore. (Questo è ovviamente vero per quanto riguarda i sistemi di media complessità, cioè le macchine non banali: l'evoluzione culturale è così rapida che lo si può constatare in prima persona.)

Ogni sistema fisico può anche essere studiato indipendentemente dal posto che occupa nella gerarchia di complessità organizzativa: così considerato, il sistema si rivela descrivibile in molti modi, che dipendono essenzialmente dalle sostanze che lo compongono e dai processi che si ritiene avvengano in esso. In generale, ci sembra corretto affermare che di fronte ad un sistema ci si possono porre le seguenti domande:

- Come si comporta nel suo insieme?
- Quante e quali parti lo costituiscono?
- Come interagiscono tra loro le parti?
- Quali fenomeni si possono associare al funzionamento del sistema?

Circa queste domande, si impongono due osservazioni:

1) A seconda dei casi, può essere disponibile più di un linguaggio per descrivere il medesimo sistema e le nostre risposte alle domande suddette sono fortemente influenzate - come sempre accade - dal linguaggio utilizzato.

2) L'esistenza di tali domande (sul sistema nel suo insieme, sulle sue parti e sui fenomeni ad esso associati) indica un vero e proprio principio epistemologico: *il principio della descrivibilità dei sistemi a più di un livello di analisi*. Da un punto di vista emergentistico, nel caso dei sistemi complessi (biologici), si possono specificare tre livelli di analisi: quello del sistema come totalità, cioè il livello olistico (o sistemico o globale); quello delle parti e della loro organizzazione, cioè il livello mereologico (o meristico); quello dei fenomeni emergenti: la vita in quanto tale, la mente, la società in quanto tale.

E' interessante notare che, nel caso dei sistemi biologici più semplici, la precisazione dei tre livelli di analisi è incredibilmente recente. Possiamo dire che solo a partire dai lavori di biologia teorica di Maturana e Varela è stato possibile

distinguere, da un punto di vista scientifico, i livelli di analisi olistico e meristico da quello della vita in quanto tale. Consideriamo, per chiarire questa distinzione, una cellula vivente.

Il livello di analisi olistico di tale sistema (la cellula come totalità) è quello che utilizzano i biologi quando studiano la cellula nel dominio delle sue *interazioni con ciò che è esterno* all'organizzazione cellulare. A questo livello, dimensione che Maturana e Varela chiamano *allopoietica*, la cellula viene studiata in termini di input e output, ovvero in termini di perturbazioni provenienti da ciò con cui la cellula interagisce e di cambiamenti interni che compensano le perturbazioni. Il livello di analisi meristico riguarda invece le parti della cellula, cioè quelle che Maturana e Varela chiamano le *sottomacchine allopoietiche* [Maturana - Varela 1973, 135].

Questi due livelli di analisi sono del tutto tradizionali. E' un merito di Maturana e Varela aver specificato un terzo livello, che non coincide con nessuno degli altri due. E' questo il livello della *vita in quanto tale*, cioè la dimensione di analisi che i due autori chiamano *autopoietica*. A questo livello interviene una descrizione completamente diversa dalle precedenti, che non fa riferimento a input o output né ad *alcunché di esterno* all'organizzazione cellulare. L'applicabilità alla vita in quanto tale e alla mente del medesimo concetto di emergenza, che è una delle tesi-cardine dell'emergentismo, sarebbe stata gravemente compromessa, almeno nella terza cultura, senza il contributo teorico di Maturana e Varela.

Nel caso del sistema nervoso - al contrario di quanto è accaduto per gli altri sistemi biologici - la maggiore difficoltà epistemologica non è stata *distinguere* i tre livelli di analisi, bensì *avvicinarli*. Consideriamo i primi due livelli: quello olistico è quello che permette ai ricercatori di studiare il sistema nervoso nel suo insieme (e, così studiato, esso risulta una meravigliosa "centrale di controllo"); il livello di analisi meristico è invece il livello utilizzato dai neurobiologi che studiano i neuroni e le loro forme organizzative (sinapsi, reti, colonne, mappe topografiche, mappe rientranti, ecc.). Benché non coincidenti, questi livelli di analisi riguardano i medesimi oggetti fisici e i medesimi processi fisiologici.

Il terzo livello di analisi del sistema nervoso, quello *psicologico*, è qualitativamente diversissimo. Questo livello è così ben distinto dagli altri due che, non appena si tenta un avvicinamento tra i tre, nascono problemi di vastissima portata,

proprio come il problema mente-corpo. *Tutti i paradigmi filosofici, senza alcuna eccezione, sostengono qualche forma di dualismo che rispecchia la radicale distinzione tra il livello di analisi psicologico e gli altri due livelli.*

Tali forme di dualismo sono: per l'approccio "soggettivistico-personologico", il dualismo persona/cervello; per il funzionalismo, il dualismo della teoria computazionale della mente (software/hardware); per i fisicalisti non-riduzionisti, il dualismo tra le proprietà sopravvenienti e quelle fisiche; per gli eliminativisti forti verso la mente, il dualismo tra organismo e comportamento; ecc.

Naturalmente i singoli autori hanno spesso personalizzato la forma di dualismo preferita; ma *tutti, nessuno escluso*, ne sostengono qualcuna. Tra queste forme personalizzate di dualismo, abbiamo visto quella di Popper ed Eccles: il dualismo sostanziale (o "quasi-sostanziale") tra Mondo 1 e Mondo 2; quella di von Hayek: il "dualismo pratico" dovuto alla nostra limitazione cognitiva; quella di Dennett (e di tanti altri): il dualismo tra macchina virtuale seriale e cervello con architettura parallela. Ma anche gli autori che non abbiamo incontrato propongono versioni personalizzate di dualismo (celebre, per esempio, quella dello psicologo Ray Jackendoff: il dualismo tra mente computazionale e mente fenomenologica). La domanda, a questo punto, è inevitabile: e gli emergentisti?

L'emergentismo è ben lungi dal costituire un'eccezione: nel concetto stesso di emergenza è presente il riferimento a *due* piani che il soggetto epistemico utilizza per distinguere il sistema studiato congiuntamente ai fenomeni emergenti, dal sistema studiato separatamente da essi. *Due piani significa dualismo*. Dualismo, però, non significa necessariamente "ontologia cartesiana", cioè dualismo di sostanze. I due piani non sono piani ontologici. *Allora* sono piani epistemologici? Questa domanda è tendenziosa, perché presuppone che qualunque dualismo debba essere un dualismo *o di sostanze o di modalità descrittive di una medesima sostanza*. Chiariamo meglio.

I due piani di cui alla definizione di emergenza corrispondono semplicemente alla differenza tra il livello di analisi psicologico e gli altri due livelli: un cervello in funzione può essere analizzato come sistema materiale (in quanto globalità o come insieme organizzato di parti) oppure considerando i fenomeni emergenti ad esso associati. E questi ultimi non sono di natura sostanziale, ma processuale. Non ha quindi senso chiedersi se si tratta di piani ontologici o epistemologici. *Se l'emergenza*

fosse sostanziale, *allora* sarebbero piani ontologici: ma i fenomeni emergenti *non sono considerati sostanze*. D'altra parte, potremmo dire che si tratta di due piani epistemologici - e non ontologici - solo se sostenessimo che nel mondo esistono solo sostanze (e *nessun* fenomeno di natura processuale): in questo caso, rifiutando alla mente la qualifica di sostanza, la ridurremmo effettivamente ad una modalità descrittiva *del cervello*. Ma non è questo ciò che l'emergentismo sostiene.

I fenomeni emergenti sono *sui generis*: non sono *né sostanze* (e quindi il dualismo degli emergentisti non è ontologico), *né descrizioni* di sostanze (e quindi il dualismo degli emergentisti non è epistemologico). A questo punto ci si potrebbe chiedere: «ma allora, che cosa sono?», ma chiederselo significherebbe *rifiutare l'emergenza come categoria di fenomeni sui generis* - il che ci riporterebbe al problema della ruota tibetana (cfr. sopra).

Conclusioni sul concetto di emergenza

Un'altra conclusione a cui siamo pervenuti alla fine della nostra ricerca è che il concetto di emergenza ha raggiunto la notorietà, almeno in parte, in seguito ad un processo che potremmo definire “ping-pong legittimatorio”. In che cosa consiste questo processo?

Consiste nell'usare una parola dal suono particolarmente evocativo in punti importanti della propria argomentazione, facendo riferimento ad altri per quanto riguarda la sua definizione rigorosa; se questo affidarsi ad altri diventa un'abitudine, può capitare che una parola venga usata da molti autori senza *mai* ricevere una definizione precisa - anzi, più viene usata, più appare “legittimo” usarla senza darne una definizione. Questo è il “ping-pong legittimatorio”. Secondo chi scrive, la parola «emergenza» è stata protagonista di un simile processo. E' frequente, infatti, incontrare in letteratura espressioni come: «i cosiddetti “fenomeni emergenti”»; ma un fenomeno, per essere *cosiddetto* “qualcosa”, deve essere *detto tale da qualcuno*. Se nessuno lo dice (cioè lo definisce) tale, *non è cosiddetto*.

E' per questa ragione che, nel corso della tesi, abbiamo proposto una nostra definizione di emergenza a cui abbiamo anche fatto costante riferimento: perché ci

sembra che l'unico modo per interrompere i "ping-pong legittimatori" sia proporre *effettivamente* delle definizioni. Questo non significa togliere importanza al potere evocativo di certi termini: tuttavia, chi scrive ha il sospetto che spesso un uso *puramente* evocativo delle parole nasconda delle profonde lacune teoriche. Le parole evocative, così come le metafore, sulla efficacia delle quali è superfluo soffermarsi, dovrebbero *aggiungersi* - e *non sostituirsi* - alle teorie positive.

In verità, quella che è stata proposta non è una definizione, bensì *una classe di definizioni*, della quale ciascun membro è caratterizzato da una differente interpretazione dell'*imprevedibilità* delle proprietà emergenti. In questa sede intendiamo proporre una nostra interpretazione dell'imprevedibilità, caratterizzata dal fatto che, con essa, torniamo - in parte - alle due dimensioni del reale (si ricordi che la nostra definizione "univa" il punto di vista diacronico e quello sincronico).

La nostra interpretazione è la seguente: *le proprietà dei fenomeni emergenti sono imprevedibili sia dal punto di vista diacronico, che da quello sincronico; non solo: esse sono imprevedibili, da entrambi i punti di vista, per molte ragioni differenti*. I motivi per cui ci siamo convinti che l'imprevedibilità vada interpretata in questo modo sono facilmente riassumibili.

Per quanto riguarda la dimensione *diacronica* del reale, nel corso della ricerca abbiamo visto che l'imprevedibilità dei fenomeni naturali è stata associata, dai neodarwinisti, alla *contingenza* e, dagli scienziati della complessità, al carattere statistico di certe leggi *necessarie* e invarianti; inoltre, essa può anche essere associata ai limiti *di fatto* dei modelli scientifici reali. In verità, *non abbiamo incontrato alcuna ragione per adottare una sola di queste interpretazioni ed escludere le altre*: anzi, crediamo di poter affermare che molti dei fattori che, anche singolarmente considerati, rendono imprevedibili i fenomeni emergenti *sono all'opera contemporaneamente*. Nel corso dell'evoluzione, infatti, agiscono probabilmente:

- casualità sotto forma di indeterminazione (per esempio: le mutazioni genetiche);
- casualità sotto forma di accidentalità (per esempio: le estinzioni di massa);
- "ex-attamenti" (cioè i cambi di funzione);
- fenomeni di auto-organizzazione (per esempio: l'origine della vita).

Riteniamo che ciascuno di questi fattori sia, anche da solo, più che sufficiente a far sì che molte delle classi di fenomeni descritti dagli attuali modelli dell'evoluzione siano decisamente imprevedibili. In questo senso *ogni specie naturale è imprevedibile e quindi, con essa, sono imprevedibili tutte le sue peculiarità*. In particolare, la nostra mente va considerata imprevedibile *almeno* tanto quanto lo è la specie *Homo sapiens*.

D'altra parte, per quanto riguarda la dimensione *sincronica* del reale, sono almeno due i fattori che rendono imprevedibili i fenomeni emergenti. Il primo fattore è legato al fatto che le parti dei sistemi ai quali associamo le emergenze interagiscono in modo complesso. Come si ricorderà, Bechtel e Richardson rispondono alla domanda «di che cosa l'emergenza è una conseguenza?» proprio con la risposta: *delle interazioni complesse tra le parti del sistema*. I due autori, dicendo questo, hanno in mente interazioni che, nel modello che le descrive, coinvolgono molti passi e passi non-lineari. Tuttavia, se i modelli stocastici dei connessionisti e i modelli del FARG sono corretti, allora le interazioni che producono i fenomeni emergenti sono complesse anche in un altro senso: esse sono *stocastiche*. In altre parole, attività mentali come il riconoscimento percettivo, la soluzione di problemi vincolati e la percezione di alto livello, nonché le azioni di massa di una società, sono rese possibili dal fatto che il risultato di ciascuna interazione tra gruppi di neuroni o gruppi di individui è puramente *probabilistico*. Se questo è vero, l'imprevedibilità di alcuni fenomeni emergenti è *anche* dovuta all'impossibilità (per ragioni di principio) di prevedere dettagliatamente le conseguenze delle interazioni tra le parti del sistema di base.

Un secondo fattore che rende imprevedibili, dal punto di vista sincronico, le proprietà dei fenomeni emergenti è che non sappiamo quasi nulla sui *vincoli organizzativi* (quelli che agiscono dal tutto alle parti). Per mettere a fuoco questa lacuna, consideriamo le differenze tra il cervello e, per esempio, il fegato. La differenza *cruciale*, quella responsabile del fatto che *il cervello pensa e il fegato no*, è quasi certamente legata alle organizzazioni funzionali dei due sistemi. Se conoscessimo le leggi delle organizzazioni, cioè se avessimo delle vere e proprie teorie sulle *organizzazioni in quanto tali*, allora - forse - potremmo stabilire *a priori* le assai diverse proprietà di fegato e cervello; ma allo stato attuale delle nostre

conoscenze non vi è assolutamente nulla che ci permetta di prevedere che da un cervello in funzione - *ma non da un fegato* - emerge il pensiero.

La nostra ignoranza delle leggi delle organizzazioni rende l'imprevedibilità della mente, a partire dal cervello, del tutto analoga all'imprevedibilità del significato, a partire dai segni di una lingua sconosciuta. In breve, ci manca ancora l'equivalente della "pietra di Rosetta" che permise di decifrare i geroglifici dell'antica lingua scritta egizia. Chi scrive è peraltro convinto che il neurobiologo Steven Rose sia nel giusto quando sostiene che *la memoria potrebbe «essere la pietra di Rosetta del cervello»* [Rose 1992, 17].

In altre parole: le leggi dei vincoli organizzativi cerebrali *potrebbero coincidere* con le leggi della memorizzazione e del ricordo biologici. Scoprire queste ultime potrebbe anche fare luce sulla natura dell'io, *in qualche modo* certamente legata alla memoria. Purtroppo, attualmente, non conosciamo ancora i segreti della memoria biologica [cfr. Rose 1992, 385].

Per tutte le ragioni qui esposte, riteniamo che i fenomeni emergenti siano imprevedibili sia dal punto di vista sincronico che da quello diacronico; inoltre, riteniamo - e con ciò concludiamo - che vi sia più di un fattore, da ciascuno dei due punti di vista, responsabile di tale imprevedibilità.

A questo punto è possibile riproporre, senza ulteriori commenti, la nostra definizione di emergenza, applicata esplicitamente alla mente, integrando in essa l'interpretazione dell'imprevedibilità che abbiamo presentato:

La mente è un *fenomeno emergente* associato al sistema nervoso centrale S degli individui di (almeno) un genere animale (gli esseri umani), perché: un osservatore, già dotato di un modello neurobiologico T di S, studiando congiuntamente S e la mente ad esso associata, (a) non riscontra alcun "cambiamento fisico" non previsto da T (stesse forze, stesse sostanze); però (b) riscontra un insieme di proprietà *psicologiche*; esse sono:

- 1) *qualitativamente nuove* rispetto a quelle neurobiologiche;
- 2) *imprevedibili* sia dal punto di vista diacronico che da quello sincronico;
- 3) *indipendenti* dalle proprietà e dall'esistenza dei singoli neuroni.

E' questa, secondo chi scrive, una possibile soluzione, interessante e dignitosa, al problema della natura della mente.

Molte questioni rimangono aperte e molte piste devono ancora essere percorse prima di poter parlare di una vera e propria “teoria emergentista della mente”. Però una certa quantità di lavoro si può considerare compiuta: alcuni nodi teorici sono stati sciolti, alcune trappole filosofiche sono state rimosse dal cammino.

Concludiamo qui la nostra ricerca, con la sensazione che la strada verso la conoscenza della mente sia ancora lunga e davvero difficile. Ma sempre più affascinante.