

Fortunato T. ARECCHI

Dipartimento di Fisica
Università di Firenze

e

Istituto Nazionale di Ottica
L.go E. Fermi, 6, 50125 Firenze

COMPLESSITÀ E DINAMICA DELLA SCOPERTA SCIENTIFICA

Conferenza tenuta il giorno 27 Aprile 1998

La scienza dei calcolatori ha elaborato una nozione formale di complessità, misurandola con la quantità di risorse (spazio di memoria, tempo di esecuzione) necessarie per risolvere un dato problema, portandoci a evidenziare due limiti dei modelli matematici: uno formale legato al teorema di indecidibilità di Gödel e uno pratico, di intrattabilità, legato alla crescita esponenziale del tempo di calcolo con il numero di variabili del problema.

Peraltro, l'esplorazione di fenomeni affetti da caos deterministico, o di sistemi estesi che si strutturano in domini spaziali, ha mostrato come in genere la competizione fra molti individui porti a situazioni macroscopiche che ammettono una gerarchia di livelli di descrizione. Chiamando complessità la coesistenza di livelli di descrizione non mutuamente riducibili, si mostra come le situazioni più rilevanti del mondo che ci circonda siano complesse.

Ne emerge una conseguenza epistemologica: la scoperta scientifica non è una procedura univoca, automatizzabile in un programma di calcolo. Qualunque modello apriori (collezione di ipotesi) va

confermato e arricchito in corso d'opera con dati sperimentali. La conferma non è una semplice procedura Bayesiana, in quanto i dati probatori delle ipotesi sono a loro volta acquisiti con apparati di misura basati su quelle ipotesi. Si fuoriesce da un argomento circolare solo rinunciando a un punto di vista unico (cioè a un'unica classe di ipotesi) e armonizzando in maniera adattiva risultati conseguiti da punti di vista diversi.

Questa adattività presuppone delle scelte che non sono deducibili dal modello in quanto ne rappresentano la condizione. E' questo il problema della creatività in qualunque impresa umana.

Si mostrano esempi di riconoscimento e controllo adattivo di dinamiche caotiche, mostrando come l'adattività eviti l'intrattabilità.

1 La sfida della complessità

Il cercare una "costruzione logica del mondo" a partire da un corpo ristretto di assiomi non solo si blocca su situazioni indecidibili in senso formale, cioè su predicati di cui gli assiomi non sono in grado di stabilire se sono veri o falsi (Svozil), ma porta a problemi intrattabili nel senso del calcolo.

L'intrattabilità può essere spiegata nel modo seguente. Nell'esecuzione di un calcolo, ci troviamo di fronte a un albero di decisioni come in fig. 1, con nodi ai quali si presenta una doppia scelta. In mancanza di elementi esterni, che aiutano a decidere, entrambe le linee che emergono vanno esplorate e questo si ripete ad ogni nodo, con una crescita esponenziale del numero di stati finali.

Chiamando "biforcazione" lo sdoppiamento che avviene ad ogni nodo, vediamo che dopo una biforcazione abbiamo due canali, dopo due ne abbiamo $2^2 = 4$ e così via. Dopo 100 biforcazioni avremo $2^{100} \sim 10^{33}$ possibili opzioni. Se dovessimo esplorarle tutte, e se impiegassimo 1 milionesimo di secondo per ciascuna (ammettendo di avere un computer molto veloce) ci occorrerebbero ancora $10^{33-6} = 10^{27}$ secondi, pari a 10^{10} volte l'età dell'universo (che è 10^{10} anni, cioè 10^{17} secondi).

Analogo albero con divergenza esponenziale si presenta in dinamica dei sistemi non lineari, ed è la base della complessità.

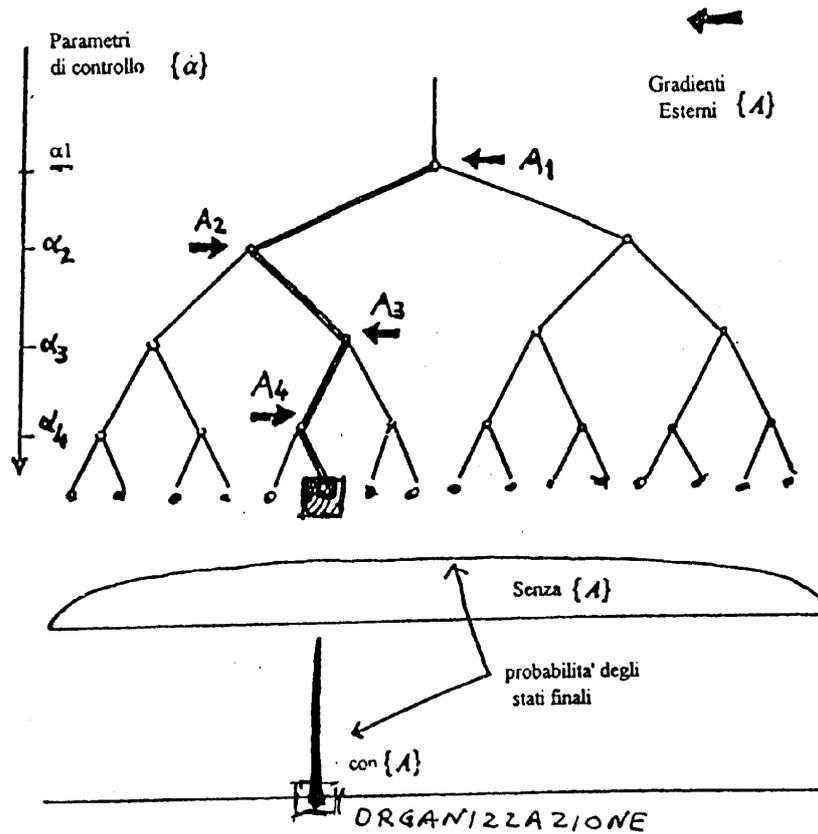


Figura 1: Albero di biforcazioni di una dinamica nonlineare al variare di un parametro di controllo α , che è "esterno" alla dinamica stessa. In basso: distribuzione di probabilità piatta degli stati finali in caso di biforcazioni simmetriche; invece l'"organizzazione" (stato finale unico) implica rottura di simmetrie dovute a forze esterne A.

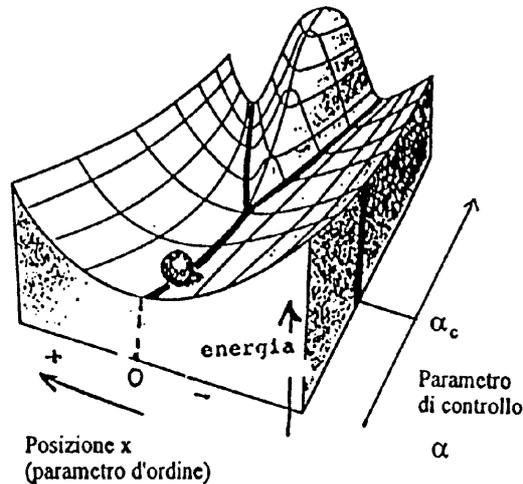


Figura 2: In base alla formulazione variazionale della dinamica, un sistema dinamico con parametro d'ordine x tende al punto di minimo di un paesaggio dell'energia la cui forma dipende da un parametro di controllo α . Se α varia, passando per il punto α_c il paesaggio ha un cambiamento qualitativo, scompare il singolo minimo e compaiono due valli laterali (biforcazione).

Un sistema dinamico con dissipazione (cioè non idealmente isolato dal resto del mondo) può essere visto come una pallina pesante che “cade” fermandosi nel punto di minima energia, detto attrattore (fig. 2 - sinistra).

Supponiamo ora che al variare di un “parametro di controllo”, α , si modifichi il paesaggio dell'energia come in fig. 2 da sinistra verso destra. Ad un certo punto (α_c) scompare la singola valle e ne compaiono due distinte: c'è stato un cambiamento “qualitativo” nella dinamica perché sono cambiati gli attrattori. Se rappresentiamo le posizioni degli attrattori in funzione di α , vediamo che quanto illustrato in fig. 2 appare come una “biforcazione” (fig. 3).

Tornando a fig. 1, possiamo ora interpretare l'albero di biforcazioni come l'evoluzione di un sistema dinamico non lineare con successive biforcazioni controllate dai valori $(\alpha_1, \alpha_2, \dots, \alpha_n) \equiv \{\alpha\}$ di un parametro di controllo (abbiamo indicato con $\{\cdot\}$ la collezione di tutti gli α).

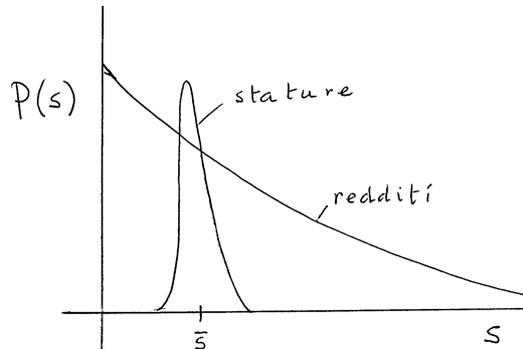


Figura 3: *Diagramma monoscala della distribuzione delle stature e diagramma a molte scale della distribuzione dei redditi.*

L'albero somiglia all'albero filogenetico di un modello di speciazione biologica. In tal caso, $\{\alpha\}$ può rappresentare un gruppo di fattori ambientali che destabilizzano una specie e la sostituiscono con nuove specie più stabili.

Se ogni biforcazione fosse simmetrica, tutti gli stati finali sarebbero ugualmente probabili, e il tempo di esplorazione crescerebbe esponenzialmente con l'ordine della biforcazione. Diremo che tale descrizione modellistica è *intrattabile*, o *complessa*, quando il tempo di calcolo diventa più lungo del tempo di evoluzione del sistema fisico, e pertanto si perde il vantaggio della predizione.

Quando però vi sia un gradiente esterno, allora la biforcazione è imperfetta e si ha una probabilità altissima di un percorso non ambiguo (linea a tratto doppio di fig. 1). Nella figura abbiamo indicato come la sintassi $\{\alpha\}$ venga modificata dagli agenti esterni

$$(A_1, A_2, \dots, A_n) \equiv \{A\}$$

che forzano il sistema ad assumere un solo stato finale (per convenzione indichiamo la successione ordinata degli A con la parentesi graffa, come abbiamo già fatto con gli α). Diremo allora che il sistema si è *organizzato*. Nell'ambito del modello rappresentato dall'albero di fig. 1, con una sintassi riassunta dal gruppo $\{\alpha\}$ di parametri di controllo, se vogliamo spiegare l'organizzazione, dobbiamo valutare ad ogni nodo il corrispondente disturbo esterno A che ha sbilanciato

la biforcazione. Andando a ritroso, ricostruiamo tutto il pacchetto $\{A\}$ degli agenti esterni che hanno modificato la sintassi.

Dunque la verità non è limitata alla “certa” applicazione delle regole $\{\alpha\}$, ma richiede anche di prendere atto degli elementi esterni $\{A\}$, cioè la verità è *regole interne più agenti esterni* (sintassi più semantica).

$$\{\alpha, A\}$$

Ma allora la verità sembra dipendere dal modello scelto. In effetti, possiamo descrivere lo stesso stato osservato nell’ambito di un modello $\{\alpha\}$, o $\{\beta\}$, e in corrispondenza dovremo considerare il gruppo di agenti esterni $\{A\}$ o $\{B\}$.

Avremo allora le descrizioni scientifiche equivalenti

$$\{\alpha, A\}, \{\beta, B\}, \dots$$

Sembrerebbe perciò che l’irruzione della realtà assuma aspetti diversi a seconda della teoria.

Affronteremo i modi di risolvere questa difficoltà nel paragrafo 3. Nel frattempo, è opportuno rimuovere una obiezione di principio mossa al problema della complessità.

La fig. 1 rappresenta un sistema “aperto”, cioè un sistema in cui i parametri di controllo α non dipendono a loro volta dalle variabili del sistema la cui informazione è racchiusa in un vettore di stato che chiamiamo x . Invece per un universo chiuso quale ipotizzato dalla teoria del tutto (TOE=Theory of Everything), qualunque parametro α è a sua volta funzione del vettore di stato. Si noti che x denota tutti i gradi di libertà del sistema, quindi se ci riferiamo all’universo conosciuto, sarà dell’ordine di tutti i nuclei esistenti, cioè dell’ordine di 10 elevato alla potenza 80.

Ora, anche se si arriverà ad una equazione del tutto, il suo valore euristico sarà scarso, in quanto noi osserviamo sempre un piccolissimo sotto-insieme y di x , e il complemento \bar{y} (tale che $y + \bar{y} = x$) ci sfugge. La dinamica di y sarà la proiezione dell’equazione del tutto nel sottospazio y . Perciò gli α includono una dipendenza dalle quantità ignote, cioè l’evoluzione di y che, sottolineo, è la sola parte del mondo a noi accessibile, sarà retta da un’equazione i cui parametri di controllo sono imposti “al di fuori” del sistema, e variano in modo non descrivibile all’interno del modello fisico del mondo

osservato. Ciò rende plausibile il modello di mondo aperto rispetto all'astrazione dell'universo chiuso, e rappresenta la sfida della complessità.

Queste considerazioni sono generali, e pertanto applicabili anche al problema di formulare frasi significative all'interno di un linguaggio (v. App. B). Ciò ci permette di formulare una dinamica della *creatività* umana, di cui la scoperta scientifica è un caso particolare (v. paragrafo 4).

2 Una risposta debole alla complessità: le leggi di scala

(v. Appendice C)

Se un effetto s è la somma di un numero grande di cause a media nulla e mutualmente scorrelate, il teorema del limite centrale assegna ad s una distribuzione di probabilità gaussiana, con media nulla e con scarto medio che rappresenta l'*unica* scala di s , nel senso che valori molto differenti da sono altamente improbabili. Ad esempio, se la statura media della popolazione adulta è 1,7 metri, sarà possibile incontrare individui alti 1,5 oppure 1,9 metri, ma sarà scarsamente probabile incontrare qualche adulto alto 0,5 oppure 3 metri.

Per contro, esiste evidenza empirica di classi di eventi di grandezza s la cui probabilità di accadimento $P(s)$ "decade" lentamente (fig. 3) e che pertanto possono essere osservati su *molte scale* con probabilità significative.

Ricordiamo alcune situazioni di questo tipo.

- Aziende, ditte commerciali oppure ricchezza individuale (Pareto, 1890): s = numero di addetti, oppure ammontare del reddito individuale;
- Frammentazione di un sasso per percosse successive (Kolmogorov, 1940): s = grandezza di un frammento;
- Parole di un testo in lingua inglese (Zipf, 1940): s = (rango = delle parole, dove il rango è definito come il numero ordinale che caratterizza la frequenza d'uso di una parola nei testi);
- Valanghe e terremoti (Bak, 1987): s = dimensione delle valanghe (o scala di un terremoto);

- Ritmo cardiaco (Stanley et al., 1991): s = deviazione del ritmo cardiaco dalla media.

In tutti questi casi, $P(s)$ è la probabilità di accadimento dell'evento di grandezza s .

In fig. 3 è rappresentata la distribuzione “monoscala” delle stature e la distribuzione “multiscala” dei redditi per la popolazione di un Paese. Non c'è da aspettarsi individui alti il doppio della media; mentre è facile aspettarsi redditi 2, o 4, o anche 10 volte più alti della media. Una particolare legge “multiscala” è la probabilità inversa $P(s) = 1/s$. Per comprenderne il significato, vediamo che aumentando s di un fattore 10 si riduce $P(s)$ al 10% che è ancora un valore facilmente osservabile.

Si dimostra che la probabilità $1/s$ corrisponde a cascate di cause indipendenti (prodotto delle probabilità) invece che a somme come nella gaussiana. Cause indipendenti, sia che si sommino sia che si condizionino con regole di prodotto, implicano delle fluttuazioni casuali nel corso dell'evoluzione dinamica (paradigma darwiniano).

Torniamo all'albero di biforcazioni di fig. 1. Considerandolo come una cascata di frammentazioni a partire da uno stato unico, la distribuzione delle probabilità finali permetterà di ricostruire eventuali correlazioni fra le diverse biforcazioni, andando oltre la legge del puro caso del paradigma darwiniano.

3 Il tempo adattivo come strategia di riconoscimento di una dinamica complessa

Sfruttiamo il carattere particolare del tempo per sottrarci al modo standard di procedere nelle osservazioni scientifiche. Il modo standard consiste nell'osservare a tempi *equidistanti*, scanditi da un orologio, e classificare le corrispondenti variazioni di posizione in uno spazio N -dimensionale degli stati (in genere $N \gg 1$) ricostruendo una geometria delle traiettorie (punti fissi degli equilibri, orbite periodiche, attrattori strani del caos deterministico, nubi stocastiche di punti nell'assenza di una legge deterministica).

Proponiamoci invece di “aprire gli occhi” solo quando si verifica una regolarità geometrica: ad esempio solo quando le differen-

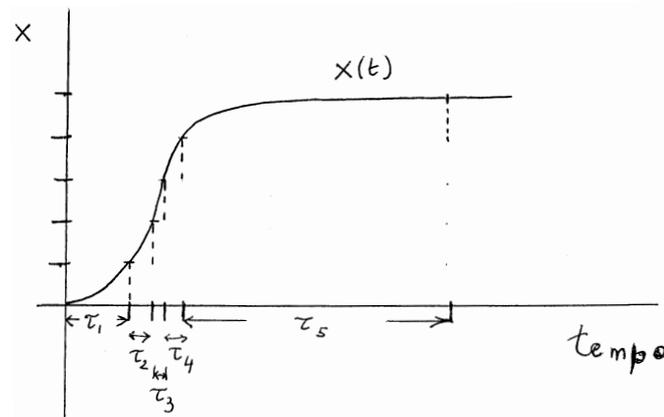


Figura 4: *Strategia adattiva: invece di osservare a tempi equispaziati e classificare variazioni di posizione, selezioniamo i tempi $\{\tau_n\}$ ai quali si minimizzano le differenze fra gli incrementi. Nell'esempio in figura, campioniamo un segnale $x(t)$ a 5 livelli verticali equispaziati. In corrispondenza, i 5 intervalli $\tau_1, \tau_2, \tau_3, \tau_4$ e τ_5 , sono molto diversi l'uno dall'altro.*

ze di posizioni fra due osservazioni consecutive (incrementi δx) si mantengono invariate (v. fig. 4) (Arecchi et al, 1994).

Ripetendo questa selezione per ciascuna delle dimensioni esplorate del sistema dinamico, e scegliendo ad ogni passo il minimo dei tempi, si ottiene una sequenza $\{\tau_i\}$ di intervalli fra istanti osservazione che è irregolare, ma ha ridotto la finestra di osservazione nello spazio ad un piccolo spessore δx che può essere meno dell'1% dell'effettiva variazione di x fra suoi valori massimo e minimo. In tal modo, invece di una informazione in N dimensioni, ci siamo ridotti a una sequenza di dati $\{\tau_i\}$ a 1 dimensione che però porta *traccia* di tutto quel che accade in N dimensioni. Stiamo applicando ad ogni passo un principio variazionale che ci fa scegliere un minimo. Se questo non è sufficiente possiamo, a un livello gerarchico superiore, introdurre un altro adattamento: in base all'informazione su come i $\{\tau_i\}$ diventano piccoli (corrispondente a geometrie rapidamente variabili) o grandi (geometrie praticamente stabili) possiamo modulare la *sensibilità* del metodo variazionale, che si mantiene come algoritmo, ma con sensibilità variabile su scale di tempi lunghe.

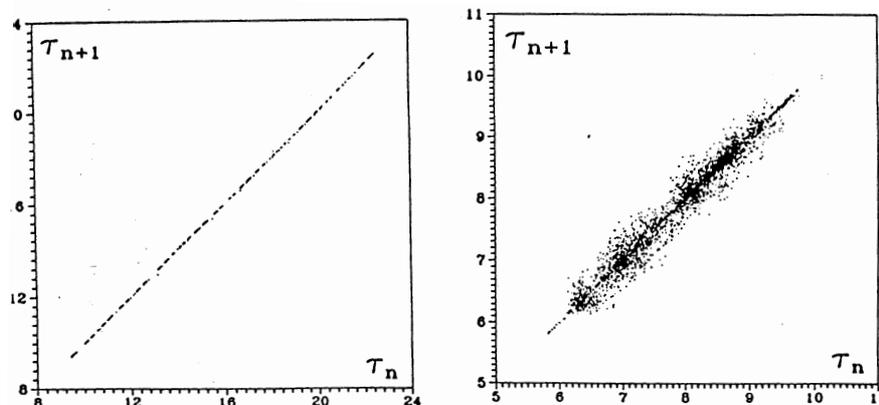


Figura 5: *Mappa di ritorno nel caso di caos deterministico (a sinistra) con forti correlazioni fra τ_{n+1} e τ_n . La lunghezza media residua attorno alla diagonale dà informazione sulla perdita di predicibilità (massimo esponente di Liapunov). A destra: perdita di correlazioni dovuta a rumore casuale aggiunto.*

Vediamo che informazione è possibile estrarre dalla sequenza $\{\tau_i\}$. Per un'orbita regolare essa si riduce a un valore unico $\tau_1 = \tau_2 = \dots$ che corrisponde al periodo dell'orbita, e l'appuntamento ad un punto geometrico particolare si riduce al metodo della sezione di Poincaré. Per un segnale affetto da caos deterministico, la sequenza $\{\tau_i\}$, pur essendo irregolare, mantiene delle correlazioni a corto raggio, cioè se l'intervallo fra l'osservazione n e $n+1$ è molto minore del tempo di imprevedibilità caotico (reciproco del massimo esponente di Liapunov Λ) avremo (fig. 5)

$$\tau_{n+1} \simeq \tau_n.$$

Di ciò ci accorgiamo disegnando la mappa di τ_{n+1} in funzione di τ_n . Vediamo che i punti si addensano attorno alla bisettrice (per cui rigorosamente $\tau_{n+1} = \tau_n$). Lo spessore residuo, mediato su tutta l'osservazione, ci dice come i τ_n si stanno decorrelando, cioè ci permette di valutare la velocità a cui il segnale ha un forte scarto rispetto alla predizione. Questa velocità, il cui reciproco rappresenta il tempo massimo T oltre il quale non ha senso fare previsioni, viene chiamata "esponente di Lyapunov della dinamica" ed indicata con la lettera Λ . Ad esempio, T è dell'ordine della settimana per eventi meteorologici

su larga scala, e dell'ordine del milione di anni per il sistema solare (che —malgrado l'ottimismo del Marchese De Laplace— è, ahimè, un sistema caotico).

Inoltre, forti deviazioni della bisettrice (forti decorrelazioni) sono indice di disturbo stocastico sovrapposto al segnale deterministico (fig. 5) e questo ci permette di estrarre la parte casuale dalla parte deterministica di un evento, e di formulare un modello dinamico appropriato.

Se adesso si disegnano mappe di ritorno (τ_{n+2}, τ_n) , (τ_{n+3}, τ_n) , ... (τ_{n+k}, τ_n) , si vede che all'aumentare di k cresce la deviazione η_k dalla bisettrice, cioè aumenta la decorrelazione dovuta al moto caotico. Ma dopo un pò si ha un'inversione di tendenza e si osservano dei minimi nel diagramma (η_k, k) (fig. 6). Ciò corrisponde al fatto che un moto caotico è decomponibile in molte orbite periodiche *instabili*: in fig. 6 noi le abbiamo individuate e possiamo stabilizzare il moto attorno ad una di esse con opportuni interventi (controllo del caos: Boccaletti e Arecchi, 1995).

Vantaggi dei tempi adattivi: i normali metodi geometrici a tempi fissi richiedono, per avere informazione utile, un numero minimo di dati n che cresce esponenzialmente con le dimensioni N del problema. Invece nel metodo adattivo la crescita è lineare; ciò vuol dire che ad esempio passando da $N = 10$ a $N = 12$, nel metodo geometrico il tempo di calcolo cresce di un fattore $10^2 = 100$, mentre nel metodo adattivo cresce di $12/10$, cioè del 20%.

Si noti che l'aver concentrato l'informazione in una stringa unidimensionale $\{\tau_i\}$ è ben più ricco di una banale proiezione su una dimensione di un moto a N dimensioni, in quanto $\{\tau_i\}$ ha esplorato *tutte* le dimensioni.

E' un pò come la strategia adattiva dei viventi. I nostri antenati nella savana non avevano bisogno di esplorare tutti gli aspetti a 3 dimensioni dei possibili percorsi nella savana, ma per sopravvivere dovevano individuare il sentiero a 1 dimensione che gli permetteva di trovare cibo e rifugio dai pericoli.

Ciò equivale a "semplificare" la complessità: invece di esplorarne tutti gli aspetti, (come in fondo fa ancora il metodo multiscale) si cerca *opportunisticamente* il percorso più conveniente per una descrizione concettuale.

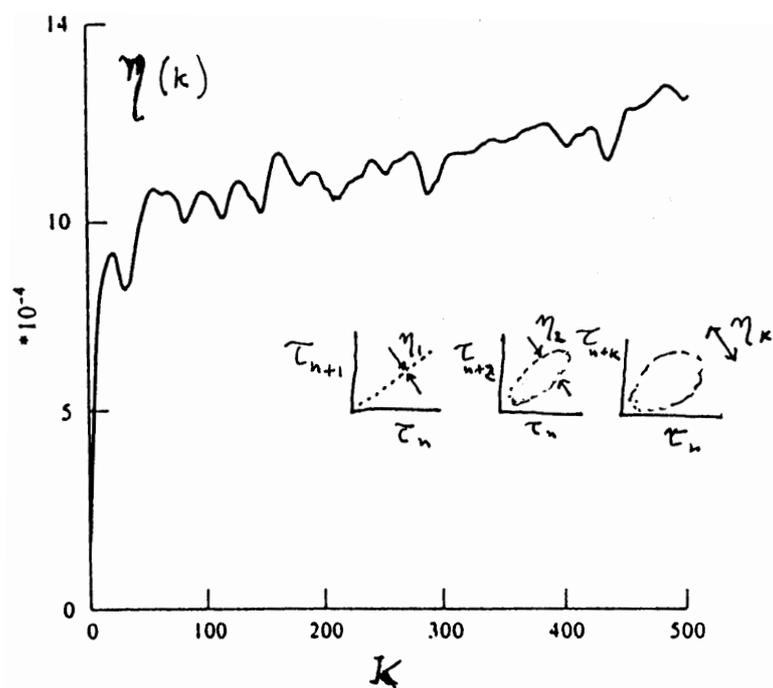


Figura 6: *Analisi adattiva di un moto caotico. Costruendo le mappe di ritorno, l'allargamento η_k attorno alla diagonale cresce con k , ma non in modo monotono. I vari minimi rappresentano le UPO's (orbite periodiche instabili) che compongono l'attrattore caotico.*

4 Conclusioni: creatività del computer e creatività umana

Un compito deduttivo a partire da un sistema di regole assegnato è eseguibile da un calcolatore. Se la scienza si riducesse a deduzioni a partire da un corpo di assiomi, allora essa sarebbe eseguibile da un computer. Ma la creatività consiste nell'innovare, cioè nel trovare concetti e relazioni nuove in presenza di situazioni osservate che non corrispondono alle nostre aspettative (v. App. A).

Johnson-Laird ha esplorato le modalità con cui un computer può essere creativo (Johnson-Laird), classificandole come segue

1. Algoritmi neo-darwiniani e genetici

Sono formati da due stadi: uno stadio *generativo*, nel quale le idee si formano in modo completamente arbitrario, e uno stadio *valutativo* nel quale le idee vengono commisurate a dei criteri: quel che sopravvive è il risultato del processo.

Se la sequenza generazione a caso - valutazione fosse fatta una sola volta, si arriverebbe a poco. Ripetendola molte volte in modo *ricorsivo*, cioè facendo tesoro dello stadio precedente, la procedura è simulata dall'algoritmo genetico di J. Holland.

2. Algoritmi neo-lamarckiani

La conoscenza può guidare la generazione di nuove idee, cioè i criteri di scelta possono non essere limitati al momento valutativo, ma intervenire già al momento formativo.

Il processo può aver luogo allora in due soli stadi senza bisogno di ricorsione.

3. Algoritmi multistadio

Le idee vengono generate sotto la guida di un criterio ma la loro forma iniziale può essere imperfetta; l'individuo le valuta alla luce di ulteriori criteri e così le idee vengono riviste e *ricombinate*. La procedura somiglia all'algoritmo genetico in quanto è ricorsiva, ma ne differisce perché i criteri di scelta sono usati sia nel momento generativo sia in quello valutativo. Questo sembra il modo più comune di procedere. Ma perché si spreca tempo a formulare un prodotto inadeguato? Ciò è dovuto al

fatto che la gente critica meglio di quanto crei, in quanto l'architettura della mente è basata su una gerarchia di processori che operano in parallelo, ma nulla sanno delle altrui operazioni e rappresentazioni interne.

Al cuore delle creatività vi sono procedure inconsce, che occorre imparare ad assemblare: donde il periodo di apprendistato che precede il padroneggiare qualunque disciplina.

Dunque, poiché i criteri valutativi possono essere comunicati a parole, mentre le capacità generative sono inconsce, i nostri giudizi critici tendono a prevalere sulle capacità immaginative.

In ogni caso, Johnson Laird ha ritenuto che un "bricolage" fra elementi di informazione già immagazzinati, più una selezione per ottimizzare la risposta, permetta a un computer di essere creativo.

Quello che egli non ha considerato è che i dati per cui si formulano delle ipotesi si riferiscono a stimoli *determinati*, quindi a un punto di vista particolare da cui si è deciso di osservare il mondo.

Al variare degli stimoli del mondo esterno, prima ancora di ricorrere a un bagaglio di ipotesi interpretative, l'uomo può variare il *punto di vista*. Ciò comporta visioni del mondo raccolte con sensibilità diverse. Si tratta di una creatività *pre-linguistica* che manca al computer, fa parte cioè dei nostri processi pre-formalizzati il costruirci scenari del mondo *non* ripetuti, ma corrispondenti a diversi punti di vista. E' qualcosa ben in più del multistadio di un computer, che ricombina elementi già archiviati in memoria.

Nel caso umano la parte pre-linguistica rappresenta l'aspetto veramente creativo che supera la macchina. Se ci limitiamo a cercare la certezza in un universo formalizzato ci blocchiamo in problemi di decidibilità Gödeliana. Se cerchiamo di attribuire un valore semantico ai nostri enunciati, ci si pone il problema della corrispondenza fra l'insieme dei nostri eventi mentali e l'insieme degli eventi del mondo: chi assicura il buon ordinamento?

Ma se consideriamo la verità come "adaequatio intellectus et rei", cioè come un programma asintotico che presuppone un continuo riaggiustamento, e questo è fatto, prima ancora di formulare nuove *ipotesi*, raccogliendo nuovi *dati*, cioè variando il punto di osservazione, ciò equivale a dire che conosciamo non solo col cervello ma con

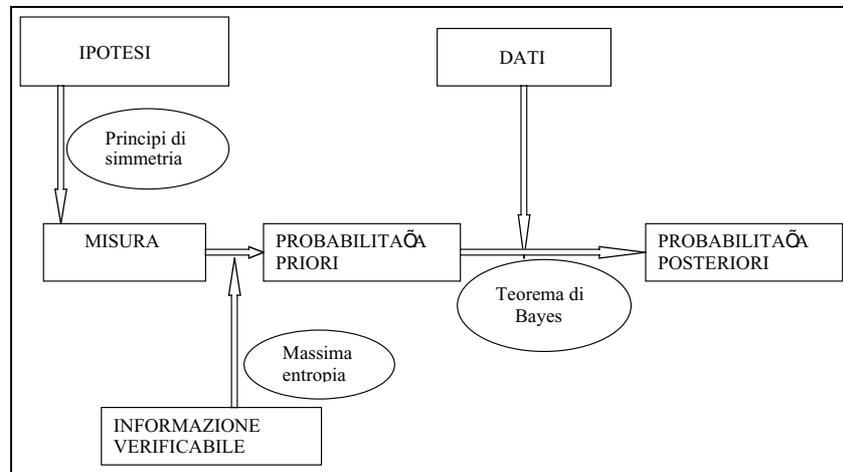


Figura 7: *Procedura bayesiana per la formulazione di una teoria.*

tutto il corpo. Questa corporeità, che rappresenta la nostra immersione nel mondo, dà un carattere di novità ad ogni conoscenza. Il dato di conoscenza non è scritto pertanto già in una Biblioteca di Babele (come nella metafora di Borges), ma viene all'esistenza quando ciascuno di noi viene a contatto con una certa parte di realtà.

Concludendo: mentre la certezza si riferisce alla correttezza di una procedura deduttiva, la verità implica una relazione fra un referente reale e un agente cognitivo disponibile ad adeguarsi cioè a ri-convertirsi rinunciando a un bagaglio di pregiudizi.

A Ruolo di ipotesi e dati nella formulazione di un modello

La fig. 7 presenta la procedura bayesiana per formulare modelli. Un ricercatore non si trova come "tabula rasa" di fronte ai dati, per due motivi:

- (i) possiede un bagaglio culturale, una memoria di situazioni pregresse, a cui attingere ipotesi
- (ii) gli stessi "dati" non piovono dall'alto, ma sono le uscite di apparati di misura che a loro volta sono costruiti in base a regole emerse da un corpo di ipotesi (è quella che gli epistemologi

chiamano una misura “theory laden”, cioè una misura carica di teoria).

Per corriamio le scatole di fig. 7. Se introduciamo la misura $m(x)$ nello spazio delle ipotesi, abbiamo introdotto un indicatore quantitativo. Se per di più possediamo della informazione corroborante, allora possiamo estrarre una probabilità apriori ($p(x)$) massimizzando l'entropia di informazione

$$- \int dx p(x) \log \frac{p(x)}{m(x)},$$

soggetta al vincolo di normalizzazione e dell'informazione verificabile. Se non c'è informazione corroborante, allora $p(x) \cup m(x)$.

In virtù del teorema di Bayes, in presenza di dati D , la probabilità aposteriori di x , $p(x|D)$ (probabilità dell'ipotesi x in presenza della condizione D) sarà

$$p(x|D) = \frac{p(x)p(D|x)}{p(D)},$$

dove $p(D|x)$ è la probabilità condizionata che si verifichi il dato D per effetto della ipotesi x (quello che faremo in sede deduttiva, deducendo D da x) e la normalizzazione è

$$p(D) = p(x)p(D|x) + p(\bar{x})p(D|\bar{x}),$$

dove \bar{x} è il complemento ad x nello spazio delle ipotesi.

Sarà:

$$p(x|D) > p(x),$$

cioè i dati sperimentali rafforzano la probabilità di un'ipotesi x .

B Polisemia del linguaggio ordinario e univocità del linguaggio scientifico*

La parola del linguaggio ordinario é polisemica, in quanto non denota un oggetto isolato ma un oggetto immerso in un contesto. Un oggetto non é mai isolabile da un contesto, ma si trova immerso nel

*Da Arecchi, 1995.

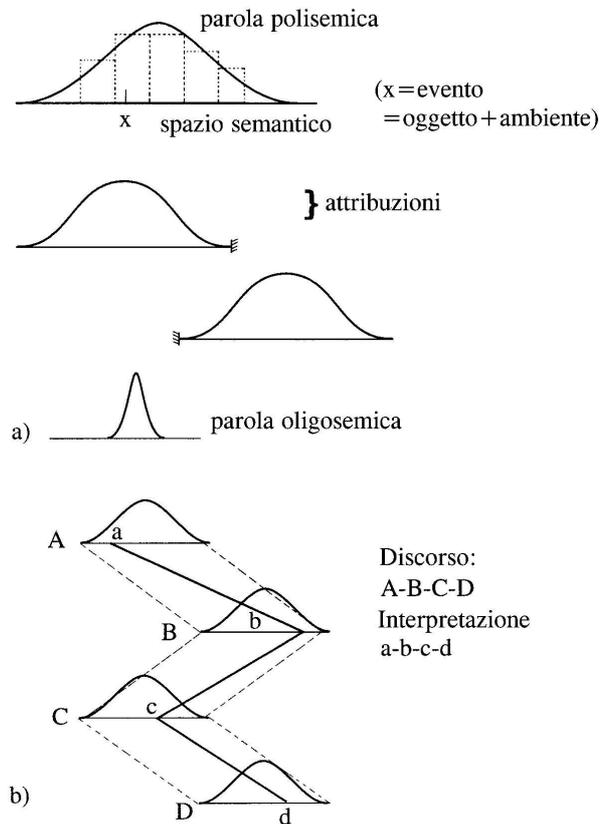


Figura 8: (a) Introdotta uno spazio semantico (asse orizzontale), ogni parola ricopre molti punti dello spazio con probabilità diverse (polisemia). Curva a istogramma: valori di frequenza attribuiti alla parola da un vocabolario della lingua. Applicando degli attributi (a loro volta polisemici), questi ritagliano dal nome un campo ristretto di significati. Il nome filtrato ha un campo semantico molto ristretto (oligosemia). (b) Sviluppo di un discorso nell'ambito del linguaggio ordinario. L'alveo delimitato da linee tratteggiate è ampio, e all'interno si possono ritagliare interpretazioni (cioè catene di correlazioni fra i termini linguistici successivi) differenti.

resto del mondo che gli attribuisce connotazioni, sfumature diverse. Chiamerò “evento” l’oggetto più il contesto. Se rappresento tutti i possibili eventi come punti di una retta orizzontale, che chiamo “spazio semantico” (fig. 8a), allora la stessa parola è usata, in modo piuttosto ambiguo, per designare tutto un ambito di eventi. La designazione può essere più o meno appropriata: questa specificità, che fa parte dell’eredità storica di una lingua, può essere vista come una distribuzione di probabilità, che è più alta là dove il termine è più pertinente. La curva si presenta come una distribuzione continua, perché le variazioni di contesto fra eventi possono essere impercettibili. In effetti, fa parte del contesto anche la nostra disposizione soggettiva a leggere un evento, e questa può cambiare in base alla nostra situazione psicologica. In un vocabolario storico questa pluralità di significati è “finitizzata” in un numero piccolo di possibili denotazioni, usate con frequenza diversa dagli autori della lingua: allora la curva continua è rimpiazzata da un istogramma.

Un discorso, visto come un flusso di parole collegate da connettivi grammaticali, appare come un vasto alveo che raccorda i vari spazi semantici e all’interno del quale si possono ritagliare interpretazioni diverse (si veda la linea in fig. 8b). Come ben si sa, un certo discorso non ammette in genere un’unica lettura, e occorre ricorrere ad elementi estranei al testo per ridurre il campo di possibili interpretazioni. Una lettura autoconsistente di un testo, in cui ogni parola sia unicamente specificata dalle correlazioni con le altre, è alquanto illusoria.

Il problema dell’interpretazione è comune a tutti i linguaggi (pittura, musica ecc.) ed è quello che impedisce il successo di una traduzione automatica da una lingua all’altra. Il filologo cerca di estrarre il senso più “fedele” di un testo, ricostruendo la storia culturale e psicologica dell’autore, cioè integrando il flusso di parole con altri elementi, che plausibilmente colgano di ciascuna parola il senso più vicino all’intendimento dell’autore. Anche così, ogni nuovo lettore che si accosta al testo lo “contestualizza” nel suo universo, e ne fa un’interpretazione in genere diversa da quella suggerita dal filologo.

Tutte queste ambiguità sembrano superate dal programma di Galileo, di non tentare le essenze ma contentarsi delle “affezioni quantitative”. Ciò equivale (fig. 9a) ad applicare un apparato di misura

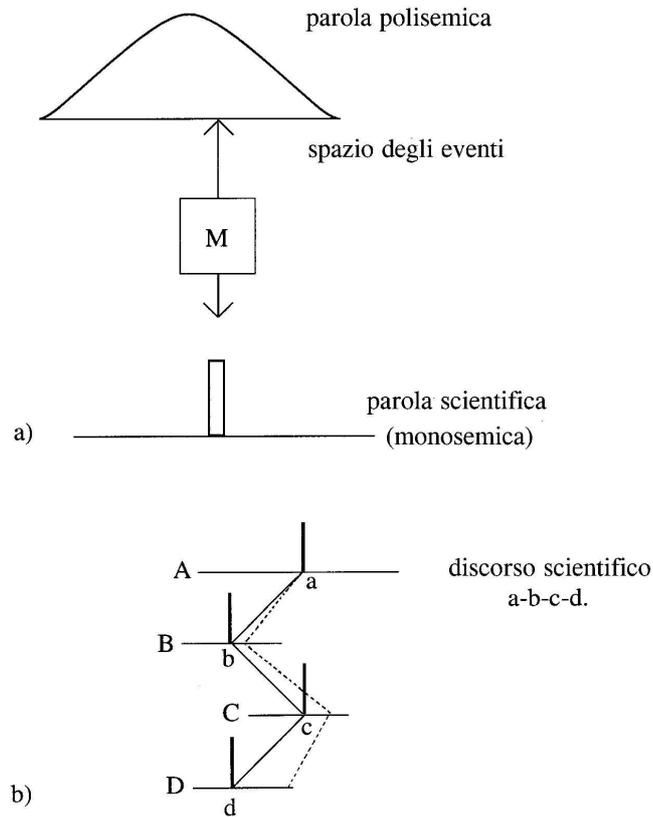


Figura 9: (a) In fisica si ottiene una univocità rigorosa “filtrando” lo spazio degli eventi con un apparato di misura M che estrae una connotazione univoca dall'evento (es. l'ago della bilancia indica solo il “peso del tavolo” invece del “tavolo”). (b) Sviluppo di un discorso nell'ambito di un linguaggio scientifico. Nei vari campi semantici (A,B,C, ...) le distribuzioni di significato estratte dall'apparato di misura M sono molto strette, e le regole di collegamento univoche (la matematica fornisce relazioni “certe” fra numeri definiti). Una piccola incertezza iniziale può però indurre notevoli variazioni per effetto del caos deterministico (linea tratteggiata).

M con una precisa regola d'uso (cioè in una precisa posizione dello spazio semantico). Il numero di uscita da M rappresenta una connotazione unica. Le nuove parole-numeri sono collegate da una nuova grammatica che è la matematica, e questa —per lo meno ai tempi di Galileo— assicura connessioni prive di ambiguità. Allora sembra che il discorso scientifico sia diventato una linea unica (linea intera di fig. 9b).

Dunque la scienza è fatta di termini univoci, collegati da regole, o leggi, necessarie. Una volta assegnato un numero sufficiente di osservazioni iniziali che permettano di stabilire sia un certo gruppo di termini sia i collegamenti fra di essi (cioè le “leggi” fisiche), si possono estrarre per via deduttiva tutte le conseguenze, anticipando eventi che hanno ancora da verificarsi: donde la predicibilità della scienza, assente nel linguaggio ordinario.

Nella scienza di Galileo, Descartes e Newton, c'è una precisa corrispondenza fra questa procedura mentale e quanto avviene nel mondo, in quanto M ha estratto le cose importanti trascurando le “qualità secondarie”, cioè quegli aspetti soggettivi del contesto che servono a fare poesia, ma non previsioni scientifiche. Allora, certezza (cioè correttezza sintattica) e verità (cioè adeguamento alla realtà) sembrano coincidere.

Questa felice armonia, che è stata la base di tutte le ideologie scientifiche, è entrata in profonda crisi per una serie di ragioni che cercherò di riassumere. Il programma galileiano di fig. 9a permette di stabilire in modo esatto se ogni individuo misurato appartiene o no a una certa classe di oggetti.

Ricordo che il “caos deterministico”, intravisto da Poincaré nel 1890 e esplorato in dettaglio da tre decenni, consiste in una estrema sensibilità alle condizioni iniziali, cioè in una instabilità delle traiettorie dinamiche, per cui da punti inizialmente molto vicini possono emergere dopo un pò traiettorie molto separate.

Se ora applichiamo allo sviluppo di un discorso le regole della dinamica nonlineare, come mostrato nella linea tratteggiata di fig. 9b, basta una piccola incertezza nell'assegnazione iniziale per portare a un discorso distinto e quindi ad introdurre l'ambiguità nel discorso scientifico. D'altronde, l'univocità di denotazione che dovrebbe essere assicurata da M non tiene conto che anche M stesso dipende da

un contesto (condizioni ambientali che possono perturbare l'apparato di misura, o variazioni nel protocollo d'uso legate alla soggettività dell'osservatore). Sembra pertanto che, in un discorso scientifico "mappabile" in regole dinamiche non lineari, tali ambiguità siano generiche, e si ripristina per lo meno, una "oligosemia" che distrugge la presunzione di univocità.

Siccome la perdita di univocità è graduale, al progredire del discorso (fig. 9b), essa può essere vista come una possibilità di scelta ai nodi di un albero di biforcazioni, quale quello di fig. 1. L'Appendice C mostra come una cascata di biforcazioni possa portare a uno sparpagliamento di probabilità su un supporto molto ampio (legge multiscala invece che monoscala: v. fig. 3).

C Distribuzione di probabilità con legge di potenza e cammini casuali di Levy

In fig. 3 abbiamo indicato la differenza qualitativa che separa una distribuzione di probabilità gaussiana (monoscala) da una distribuzione a legge di potenza (multiscala).

La gaussiana emerge come somma di segnali scorrelati (teorema del limite centrale), di conseguenza, essa rappresenta la distribuzione di probabilità di una particella sottoposta a un rumore additivo a media nulla (moto browniano).

Se invece abbiamo una legge di aggiornamento (a tempi discreti $\Delta t = 1$)

$$w(t + 1) = \lambda w(t), \tag{C.1}$$

dove λ è un rumore moltiplicativo con distribuzione a supporto finito $\Pi(\lambda)$ e con media $\langle \lambda \rangle = 1$, si vede facilmente come, passando a una trasformata logaritmica $u = \log w$, $\mu = \log \lambda$, si abbia

$$u(t + 1) = u(t) + \bar{\mu} + \delta\mu. \tag{C.2}$$

Se $\langle \lambda \rangle = 1$ allora $\bar{\mu} = \langle \log \lambda \rangle < 0$ e $\delta\mu$ sarà il rumore residuo a media nulla. Se w è sottoposto a un vincolo: $\langle \lambda \rangle = \text{finito}$, $w > w_0$, la (C.2) ammette come soluzione stazionaria la cosiddetta distribuzione *barometrica*,

$$P(u) = e^{\frac{\mu}{b}u} = e^{-\frac{1}{T}u}$$

dove D è l'ampiezza di autocorrelazione del rumore $\langle \delta\mu^2 \rangle = D$ e T è una temperatura equivalente. Tornando a w , si avrà

$$P(w) = P(u) \frac{du}{dw} = w^{-1-\frac{1}{T}} \quad (\text{C.3})$$

e imponendo i vincoli

$$\int_{w_0}^{\infty} dw w^{-1-\frac{1}{T}} = N \quad (\text{C.4})$$

si ricava $T = 1 - w_0$, cioè si esprime la temperatura mediante la troncatura inferiore. Se la fig. 3 si riferisce ai redditi, allora w_0 rappresenta un minimo di sussistenza (quale sarebbe un sussidio di disoccupazione).

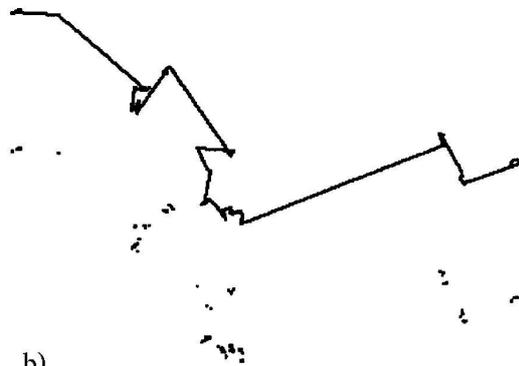
Se ora consideriamo un moto a caso (random walk) e decorreliamo un salto dal precedente, siamo nella situazione Gaussiana e i punti visitati si concentrano densamente su una macchia di raggio quadratico $\langle R^2 \rangle = Dt$, dove D è un coefficiente di diffusione (fig. 10a). Se invece consideriamo un moto a caso in cui $l > 1$ passi successivi possono essere correlati, allora otteniamo i percorsi di Levy (Levy 1927) con salti anche molto ampi che danno luogo a una intermittenza (fig. 10b). Il corrispondente insieme di punti visitati ha dimensioni frazionarie, collegate all'esponente $\alpha = 1 + 1/T$ della legge (C.3).

Salti di Levy sono osservati nella regolarità della successione di basi del DNA, nei mercati finanziari, nella turbolenza (Mallamace e Stanley, 1997). Questa universalità di comportamento, indipendente dal tipo particolare di percorso considerato, è legata alla struttura gerarchica che risulta dalla iterazione dell'equazione ricorsiva (C.1), che può essere vista come la generatrice di un albero quale adombrato in fig. 1.

Quello che succede in economia o fisiologia è pertanto qualitativamente vicino ai salti creativi che nel corso della scoperta scientifica l'investigatore compie in uno spazio di ipotesi, quale quello considerato in Appendice A.



a)



b)

Figura 10: (a) Moto stocastico Browniano con salti ai prossimi vicini su un reticolo quadrato. (b) Moto di Levy, con salti correlati (cioè nella stessa direzione) per un numero l , di passi successivi, con (probabilità della lunghezza l con legge di potenza).

Riferimenti bibliografici

- [1] F.T. ARECCHI, G. BASTI, S. BOCCALETTI, A. PERRONE, *Adaptive recognition of a chaotic dynamics*, Europhys. Lett., 27 (1994) 327-332.
- [2] F.T. ARECCHI, *Il linguaggio della scienza: verità e certezze*, KOS (Ospedale San Raffaele, Milano), n. 121, pp. 14-23 (1995).
- [3] F.T. ARECCHI, *Truth and certitude in the scientific language*, in F. Schweitzer (Ed.) *Self organization in complex structures. From individually to collective dynamics*, Gordon and Breach, London 1997, pp. 3-20.
- [4] F.T. ARECCHI, *Complexity, complex systems and adaptation*, in: F.T. Arecchi, A. Farini, *Lexicon of Complexity*, Firenze 1997, pp. 7-41.
- [5] F.T. ARECCHI, *Adaptive strategies for recognition, noise filtering, control, synchronization and targeting of chaos*, Chaos (AIP), 7 (1997) 621-634.
- [6] S. BOCCALETTI, F.T. ARECCHI, *Adaptive control of chaos*, Europhys. Lett., 32 (1995) 127-132.
- [7] P.N. JOHNSON-LAIRD, *Human and machine thinking*, Erlbann, Hillsdale NJ, 1993 (tr. it. Il Mulino, Bologna 1994).
- [8] P. LEVY, *Theorie de l'Addition des Variables Aleatoires*, Paris, Gauthier-Villars, 1927.
- [9] F. MALLAMACE E E.H. STANLEY EDITORI, *The Physics of Complex Systems*, Scuola E. Fermi 1996, Corso CXXXIV, Società Italiana di Fisica, Bologna, 1997.
- [10] K. SVOZIL, *Randomness and Undecidability in Physics*, World Scientific, Singapore 1993.