

## Fenomenologia della coscienza: creatività e libertà

F. Tito Arecchi

Università di Firenze e CNR-INO, Largo E.Fermi,6-50125, Firenze

e-mail: [tito.arecchi@ino.it](mailto:tito.arecchi@ino.it); homepage: [www.ino.it/home/arecchi](http://www.ino.it/home/arecchi)

### Sommario

Oggi si discute sul potere delle tecnologie e si cerca di porre un freno sul piano etico e del diritto con un *principio di responsabilità*. Prima ancora però, occorre domandarsi: è veramente la tecnica onnipotente, o ha dei limiti intrinseci? Sul piano di un automa in grado di eseguire una molteplicità di compiti, la cosiddetta Macchina Universale di Turing, può questa realizzare tutti i compiti accessibili a una mente umana?

In questo lavoro, si presentano situazioni *complesse* in cui una scelta richiede una *creatività* propria degli umani e assente nelle macchine

Si studiano due momenti distinti della cognizione umana, e cioè *apprensione* (A), in cui una percezione coerente emerge dal reclutamento di gruppi neuronali, e *giudizio* (B), in cui la memoria richiama precedenti brani di (A) codificati in un linguaggio conveniente, questi vengono confrontate con l'apprensione più recente e dal confronto consegue la formulazione di un *giudizio*.

Il primo momento (A) ha una durata attorno a 1 sec; esso appare come un presente atemporale e in cui si acquista consapevolezza di un aspetto dell'ambiente che stimola una adeguata risposta motoria; ad esso è associato un correlato neuronale sperimentabile (il cosiddetto NCC= *neural correlate of consciousness*). Esso può essere descritto come una interpretazione degli stimoli sensoriali sulla base di algoritmi già disponibili, mediante una procedura di Bayes.

Il secondo momento (B) implica il confronto di due apprensioni acquisite a tempi diversi, codificate in un dato linguaggio e richiamate dalla memoria. (B) dura circa 3 sec e richiede la *auto-coscienza*, in quanto l'agente che esprime il giudizio deve essere consapevole che al suo scrutinio sono sottoposte le due apprensioni e che a lui tocca estrarne una mutua relazione. A differenza di (A), (B) non presuppone un algoritmo, ma piuttosto ne costruisce uno nuovo attraverso una procedura di Bayes *inversa*. Questa costruzione di un nuovo algoritmo è intimamente legata con la *creatività* e la *libertà decisionale*. Queste procedure sono proprie degli uomini e non sono implementabili in una macchina.

### Parole chiave

Sincronizzazione neuronale, complessità, apprensione, giudizio, inferenza di Bayes, Bayes inverso

*Bios&Law-Catania 14-15 genn 2011*

## LA COGNIZIONE: INTRODUZIONE

Dobbiamo distinguere fra due tipi di processi cognitivi. Nel primo, uno stimolo esterno dà luogo a una percezione coerente, da cui consegue una definita reazione motoria, lo chiameremo *apprensione* (A); esso corrisponde al reclutamento di gruppi di neuroni corticali che rispondono in sincronismo. Nel secondo, una unità (A) codificata in qualche linguaggio viene richiamata dalla memoria e confrontata con una (A) più recente; il confronto porta alla formulazione di un *giudizio* (B). Ho usato i termini *apprensione* e *giudizio* per i due atti cognitivi che caratterizzano l'intelletto umano (Lonergan, 1957)

Il primo ha luogo su un intervallo di tempi attorno a 1 sec (da alcune centinaia di millisecondi fino a un massimo di 3 sec). Esso risulta dalla sincronizzazione collettiva delle oscillazioni registrate dall'elettro-encefalo-gramma (EEG) nella banda gamma (cioè attorno ai 50 Hertz), sincronizzazione che coinvolge aree corticali estese. Il risultato è una percezione coerente, in base a cui l'agente cognitivo acquista consapevolezza di un aspetto dell'ambiente che lo circonda (Rodriguez & al., 1999; Singer, 2007; Womelsdorf & Fries, 2007). La durata di (A) può essere allungata fino a 150 sec mediante tecniche di meditazione (Lutz & al., 2004).

Il meccanismo di sincronizzazione neuronale non presenta differenze rilevanti fra un soggetto umano e una animale superiore (scimmia o gatto). In effetti (A) può essere studiato con vari metodi di laboratorio, e qualora esso sia identificato con la Coscienza (impropriamente, come vedremo) si parlerà di correlati neuronal della coscienza (NCC).

I vari (A) possono essere richiamati dalla memoria e utilizzati per decisioni motorie; ciò accade nella vita quotidiana di qualunque agente cognitivo. Anche un allungamento della scala di tempi non altera quanto detto e parleremo sempre di processi (A).

Negli esseri umani (e - per quanto ne sappiamo - solo in essi) l'informazione di (A) viene codificata in un linguaggio conveniente e conservata non come una capacità motoria ma come una unità con un significato preciso. I vari blocchi linguistici possono essere confrontati per estrarre una tendenza globale nella sequenza. Questo accade su scale di tempi oltre i 3 sec; chiamiamo (B) questa operazione che presuppone la codifica degli (A) nello stesso linguaggio. La frontiera temporale fra (A) e (B) è stata esplorata da E. Pöppel in vari lavori (Pöppel, 1997, 2004, 2009).

(B) non è più un singolo processo di sincronizzazione, in quanto il confronto implica la presenza di differenze fra le varie (A) confrontate. Affinché (B) sia efficace, occorre che il soggetto cognitivo sia conscio della propria unità che persiste nell'esplorazione di differenti pacchetti linguistici; pertanto (B) non è mera *consapevolezza percettiva*, ma *auto-coscienza*. Spiegheremo (B) mediante quello che abbiamo chiamato "inferenza di Bayes inversa". Il Bayes inverso che verrà qui spiegato permette di formulare un *giudizio* (B) (Arecchi, 2010), mentre (A) con la sua sincronizzazione neuronale si limita all'*apprensione*.

La Fig. 1 mostra la sequenza temporale delle operazioni che hanno luogo nella neo-corteccia cerebrale in una percezione visiva. Ci rendiamo conto che abbiamo avuto una percezione coerente o *apprensione* quando uno stimolo sensoriale suscita una risposta adeguata (o una reazione motoria o una formulazione linguistica). Riferendoci alla percezione visiva, e esponendo il soggetto a una immagine a partire dal tempo  $t=0$ , lo stimolo visivo viene codificato in un treno di impulsi elettrici (spikes), ciascuno di durata 1 millisecondo e con mutua separazione (detta interspike interval = *ISI*) che codifica gli aspetti del segnale. Le spikes

viaggiano inalterate dalla retina alla corteccia visiva primaria *VI* con una velocità attorno a 1 m/sec, e pertanto arrivano a *VI* dopo 100 msec.(\*)

Da *VI* i segnali sono elaborati in due canali, il *WHAT = CHE COSA* (elaborazione di forme e colori) e il *WHERE = DOVE* (elaborazione di relazioni spaziali e temporali). Nella elaborazione, i segnali di partenza (detti *bottom-up* = dal basso verso l'alto) sono modificati da segnali *top-down* (= dall'alto verso il basso) che provengono dalle memorie semantiche (percezioni precedenti categorizzate, emozioni, attenzione, ecc). La modifica è una ri-codifica che mantiene l'informazione di ingresso ma la presenta sotto altra forma dopo averla arricchita con valutazioni soggettive. I due flussi separati convergono verso la corteccia pre-frontale (*PFC = pre-frontal cortex*), dove ha luogo una lunga elaborazione. Dopo un ulteriore mezzo secondo, una decisione viene trasmessa alle aree motorie o del linguaggio, e quindi possiamo essere edotti di una percezione coerente del soggetto attraverso le sue reazioni.

Ho delineato la sequenza temporale come registrata in laboratorio su soggetti normali (Rodriguez et al., 1999). La lunga elaborazione nella *PFC* corrisponde a una ricerca per tentativi con successive presentazioni *top-down* di interpretazioni alternative, scegliendo alla fine quella che meglio si adatta agli aspetti dello stimolo di ingresso. Questa procedura è stata elaborata in svariati modelli computazionale. In particolare, io trovo utile la teoria della risonanza adattiva (*ART=adaptive resonance theory*) di Grossberg (Grossberg,1987; Carpenter & Grossberg, 2003).

Se si richiede una reazione rapida (per esempio, frenare l'auto ad una emergenza o rispondere a un servizio a tennis), gli 800 msec sarebbero un tempo troppo lungo, e allora riduciamo la durata della elaborazione evitando una ricerca prolungata e ricorrendo a procedure già codificate attraverso una *strategia Bayesiana* che verrà illustrata qui di seguito.

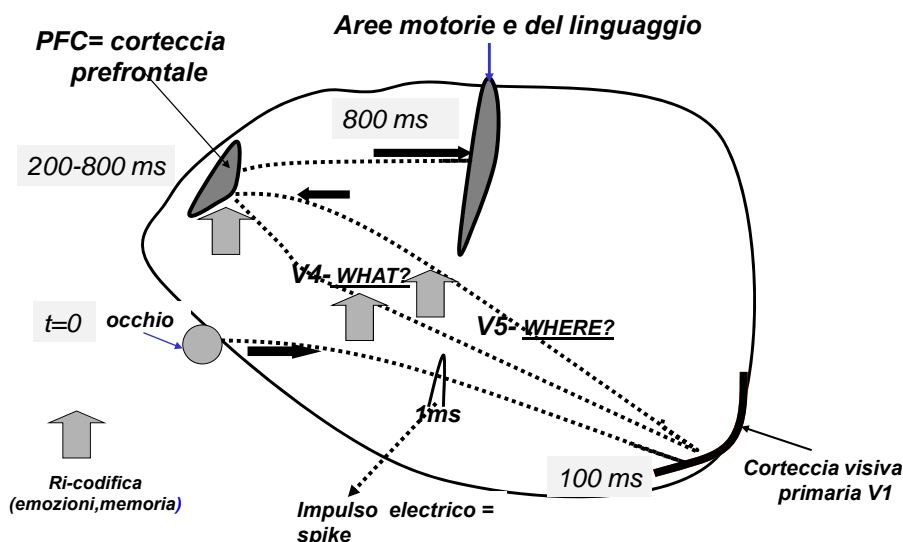


Fig.1 -Percezione visiva: percorsi e tempi. La sottile linea di contorno è una rappresentazione schematica della neo-corteccia (spessore circa 2 mm). Presentiamo la sequenza temporale degli eventi a partire dall'arrivo sulla retina di uno stimolo visivo al tempo  $t=0$ . Dopo 100 msec, il segnale codificato come un treno di impulsi elettrici raggiunge V1. Dopo V1, il segnale è elaborato in due aree distinte, V4 o WHAT, e V5 o WHERE. Le informazioni separate sono ricombinate in PFC, insieme con i segnali *top-down* che vengono dal cervello interno (emozioni, memoria). Questo mescolamento richiede circa mezzo secondo partendo da 200 msec; ad 800

*msec dopo l'applicazione alla retina, emerge una decisione che attiva le aree motorie e del linguaggio*

## IL CERVELLO CAOTICO

Per circa 800 msec ( ma arrotondiamo a 1 sec, visto che ci sono forti fluttuazioni da una percezione all'altra) abbiamo dovuto mantenere intatta l'informazione di partenza; altrimenti, i nostri resoconti sul mondo sarebbero tutti carenti. Si sa però che un singolo neurone isolato è affetto da *caso deterministico*, e dovrebbe perdere l'informazione iscritta nelle propria dinamica entro qualche millisecondo (Arecchi, 2004). E' questa la correzione al determinismo di Newton-Laplace –che cioè, note le forze, le condizioni iniziali determinano tutto il futuro – dovuta a Henri Poincaré e che oggi appunto chiamiamo *caos deterministico*, indicando con un ossimoro che il caos, o perdita di informazione, non è un fatto statistico che riguardi grandi folle, ma interviene già in problemi apparentemente semplici perché retti da poche variabili dinamiche; per l'esattezza, basta averne più di 2, da 3 in su. La ragione consiste in una *esagerata sensibilità alle condizioni iniziali*. Le condizioni iniziali sono gruppi di numeri (uno per ogni variabile del problema) , ma di ogni numero reale possediamo solo una versione troncata, la cui rappresentazione geometrica è un segmento e non un punto. Si pone allora il problema: traiettorie che partono da punti diversi del segmento si mantengono vicine o piuttosto si divaricano nel corso del tempo, sì che una trascurabile incertezza iniziale dia luogo a vistose differenze al passare del tempo? Per esplorare le conseguenze di una incertezza iniziale, applichiamo piccole perturbazioni trasversalmente alla traiettoria ideale calcolata a partire da un punto. Se il sistema è trasversalmente stabile, al passare del tempo ritornerà alla traiettoria ideale; se instabile, la nuova traiettoria dopo la perturbazione si allontanerà sempre di più da quella ideale, il cui contenuto d'informazione verrà perso. Questo accade nei modelli meteorologici, per cui bisogna aggiornare i dati ogni giorno. Nel caso di un neurone isolato, il tempo di perdita dell'informazione è di qualche millisecondo (Arecchi,2004): troppo poco rispetto alle centinaia di millisecondi necessarie per decidere.

Il caos deterministico può essere controllato aggiungendo un segnale che modifichi la stabilità trasversa senza disturbare la traiettoria longitudinale (Ott, Grebogi & Yorke,1990). Il controllo equivale a una ri-codifica del problema, che così passa –diciamo- da  $N$  variabili a  $N+m$  variabili (le  $m$  variabili aggiunte sono quelle che aumentano la stabilità traversa). Pittorescamente, visualizziamo un problema dinamico a  $N$  variabili come il moto di un punto in uno spazio a  $N$  dimensioni. Il controllo del caos equivale dunque all'immersione in uno spazio a  $N+m$  dimensioni.

Applichiamo questa strategia del controllo alla dinamica dei neuroni. Per attuare la ri-codifica, dobbiamo modificare gli stimoli *bottom-up* con le interpretazioni *top-down*. Una percezione è ben lungi dall'essere una registrazione passiva di dati esterni, come invece avviene sul dorso foto-sensibile di una video-camera.

Come fa *PFC* a scegliere fra segnali diversi che gli arrivano? L'ipotesi di *feature binding* (Singer, 2007) dice che tutti i neuroni coinvolti nell'elaborare i segnali che corrispondono alla stesso aspetto (*feature*) emettono treni sincronizzati di impulsi. La sincronizzazione collettiva è un comportamento robusto che può durare abbastanza a lungo da suscitare una decisione. Essa va considerata come il risultato di stimoli individuali a neuroni più un mutuo accoppiamento con altri neuroni; il tutto insieme si configura come un controllo del caos che allunga la durata di una informazione rilevante per un tempo sufficiente ad allertare *PFC*..

E' stato ipotizzato (Baars,1989,Dehaene & Naccache,2001) che, in presenza di uno stesso stimolo esterno (bottom-up), differenti aree neuronali sono sottoposte a sincronizzazione collettiva, ciascuna con differenti contributi top-down che corrispondono a differenti interpretazioni del segnale di ingresso . Le differenti aree sincronizzate entrano in competizione fra di loro. Il dominio sincronizzato più esteso, cioè il suo modo di leggere il segnale, dominerà e orienterà una adeguata reazione motoria.

L'evidenza sperimentale di sincronizzazione in animali di laboratorio risulta correlando i segnali catturati da microelettrodi ciascuno posizionato su un singolo assone (Singer,2007). In soggetti umani i microelettrodi sarebbero troppo invasivi; d'altronde un comune EEG non dà risoluzione sufficiente per evidenziare l'effetto. Se però si filtra dallo EEG una quasi-sinusoide entro la banda gamma (intervallo di frequenze fra 40 e 70 Hertz, caratteristico delle computazioni corticali) si può evidenziare un accordo di fase di segnali filtrati estratti da aree corticali distanti, ad esempio *VI* e *PFC*. Questo accordo evidenziale sincronizzazione dei neuroni che appartengono a quelle aree ( Rodriguez et al., 1999).

## L'APPRENSIONE COME INFERENZA DI BAYES

Fin qui, abbiamo trattato la sincronizzazione come il comportamento collettivo di una folla di neuroni, ciascuno rappresentato da un sistema dinamico caotico. In effetti, nel cervello "bagnato", cioè aperto a disturbi ambientali per effetto dei collegamenti fisiologici con tutto l'organismo, un modello dinamico per cui un risultato  $d$  discende in modo univoco da un'ipotesi  $h$  attraverso una precisa relazione funzionale appare una idealizzazione poco consona alla realtà.

Per sottrarsi al semplicismo del modello dinamico, lo rimpiazziamo con un approccio statistico basato sull'inferenza di Bayes (1763).Esso consiste nella procedura seguente

- i)Partendo da una situazione iniziale, formuliamo un ventaglio di ipotesi  $h$  ,a ciascuna delle quali assegniamo una probabilità a-priori  $P(h)$ , essendo  $P$  un numero positivo compreso fra  $1$  (certezza assoluta) e  $0$  (accadimento escluso);
- ii) Ogni  $h$ , inserita in un *modello di evoluzione*, genera dati  $d$  con la probabilità condizionata  $P(d|h)$  che  $d$  risulti da  $h$  (la nozione di probabilità condizionata si esprime con una barretta verticale fra condizionato e condizionante,e- come prima- si attribuisce alla probabilità un valore  $0 < P < 1$ ;
- iii) Effettuando una misura, si registra un particolare dato  $d$  da un insieme cui spetta una probabilità  $P(d)$ ;
- iv) La combinazione di iii)+ii) seleziona una particolare ipotesi  $h^*$ , cui spetta la più alta probabilità *a-posteriori*  $P(h^*) = P(h/d)$ .

La procedura si riassume nella formula:

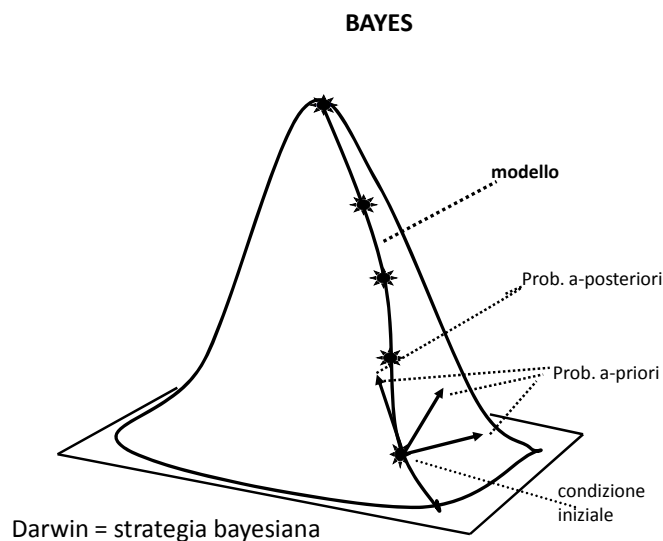
$$P(h^*) = P(h/d) = P(h) \cdot P(d|h) / P(d)$$

Dunque,  $h^*$  è l'ipotesi più plausibile, essendo quella cui spetta la probabilità più alta.

Nel corso di una indagine (pensiamo a Sherlock Holmes) abbiamo fatto un passo avanti in quanto abbiamo ristretto il ventaglio di ipotesi da  $P(h)$  a  $P(h^*)$ . Ripetiamo ricorsivamente la procedura, utilizzando lo stesso algoritmo che corrisponde al modello  $P(d|h)$  scelto. Per utilizzare una rappresentazione didascalica, riduciamo a 2 il numero di variabili del problema, in modo da muoverci su un piano, e indichiamo le probabilità con le altezze verticali. Successive applicazioni di Bayes equivalgono alla scalata di un monte di probabilità, a partire da valori bassi; il picco del monte rappresenta la massima plausibilità (Fig.2) (Arecchi,2007 a,b)

Nel Sec.XIX, la formula di Bayes ha influenzato ogni programma scientifico. L'*evoluzione* darwiniana è bayesiana, con la *mutazione* che corrisponde alla formulazione di un ventaglio di

ipotesi  $h$ ; fra queste -inserite in un modello  $P(d / h)$  di sviluppo- prevale (viene *selezionata*) quella mutazione  $h^*$  che meglio si adatta ai vincoli ambientali  $d$ .



*Fig.2-- Costruzione di ipotesi viepiù adeguate agli esperimenti, per applicazione successiva del teorema di Bayes. La procedura è una scalata del monte della probabilità guidati dal modello, o algoritmo. Le stellette indicano le posizioni successive. Ad ogni punto sulla curva che porta al picco corrisponde un valore di probabilità. Si noti che la strategia darwiniana è una implementazione di Bayes.*

Il modello  $P(d / h)$  è l'algoritmo con cui istruiamo un computer, facendone un sistema esperto (*expert system*) che seleziona le ipotesi confrontandosi con i dati.

Interpretiamo l'*apprensione* come una inferenza di Bayes (Arecchi,2007 a,b) descrivendo quel che accade entro un intervallo di sincronizzazione (attorno a 1 sec) come segue (Fig.3)..L'agente cognitivo deve scegliere l'ipotesi  $h^*$  più plausibile fra un gran numero di  $h$ . La memoria è equipaggiata con un modello procedurale  $P(d / h)$  che genera un dato  $d$  per ogni ipotesi  $h$ ; d'altra parte, l'ingresso sensoriale è un dato preciso  $d$ ; dunque la procedura di Bayes seleziona a-posteriori l'  $h^*$  che meglio si adatta al dato  $d$  (.

Il cambiare modello è una operazione di ri-codifica che ha luogo durante l'elaborazione di mezzo secondo che intercorre fra l'arrivo degli stimoli bottom-up e l'espressione di una reazione in termini di decisioni motorie. La ri-codifica top-down utilizza un insieme di modelli  $P(d / h)$  richiamati dalla memoria, scegliendo quello che i meccanismi interni (emozioni, attenzione, valori) suggeriscono come il più appropriato. Negli animali, questo insieme di modelli è costruito in precedenti fasi di addestramento, nei robot viene inserito come istruzioni; in ogni caso, l'insieme è limitato sia per gli animali sia per i robot.

Vedremo invece che un soggetto umano è capace di costruire nuovi algoritmi attraverso il confronto di brani linguistici correlati, e chiameremo *creatività* questa capacità.

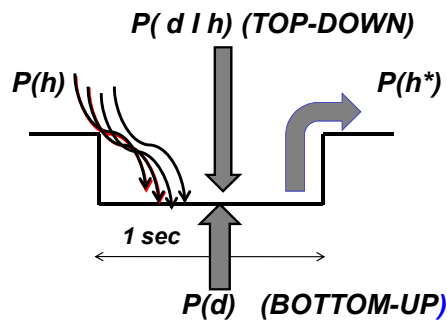


Fig.3- Apprensione come inferenza bayesiana ( $h$ =ipotesi apriori,  $d$ = dati)  
 Selezione dell'ipotesi aposteriori  $h^*$  in base a cui agire, a partire da un ventaglio di  $h$ , per azione congiunta di uno stimolo sensorio  $d$  (bottom-up) e di un modello interpretativo o algoritmo (top-down).

## LA COMPLESSITA'- IL TEOREMA DI GÖDEL

Un problema è complesso quando non è risolvibile nell'ambito di un singolo algoritmo (o di un numero piccolo e prefissato di algoritmi). La Fig.2 visualizza un processo, in cui un agente cognitivo, equipaggiato con un modello  $P(d/h)$  del mondo, interagisce in modo ricorsivo con l'ambiente, aggiornando ad ogni passo il punto di partenza. Questa strategia non funziona più in una situazione complessa. Sintetizzando tre decenni di intensi dibattiti, possiamo definire complessa una situazione che non è esaurita da un singolo modello. Lo scambio di algoritmo è una procedura non-algoritmica; un sistema complesso è visualizzabile come un paesaggio con molti monti di probabilità (Fig.4). L'arrampicarsi su un singolo versante può essere automatizzato da un programma di *massimo gradiente* (che cioè segue una linea ottimale di massima pendenza). Si tratta di una procedura *non-semiotica* (che cioè non richiede l'esplorazione dei significati). Al contrario il saltare su altri versanti, e continuare la strategia di Bayes scalando altri colli, è un atto di *creatività* che richiede una comprensione globale dell'ambiente circostante (*semiosi*) e non già una collezione ridotta di dati (riduzione del mondo a una collezione finita di simboli).

Denotiamo il paesaggio con molti colli come "*complesso semanticamente*". Ad ogni colle va attribuito un significato diverso, in quanto si deve operare su di esso con un algoritmo diverso. È stato ipotizzato (Sebeok, 1992) che la semiosi è la proprietà che distingue un vivente da una macchina universale di calcolo di Turing. Più avanti identificheremo la *creatività* con il salto da un modello bayesiano a un altro. Semiosi dunque equivale a creatività come illustrato in Fig.5. Nel campo della ricerca scientifica, la differenza fra una singola strategia bayesiana e un salto creativo è quella che è stata storicizzata come differenza fra *scienza normale* e *spostamento di paradigma* (Kuhn, 1962)

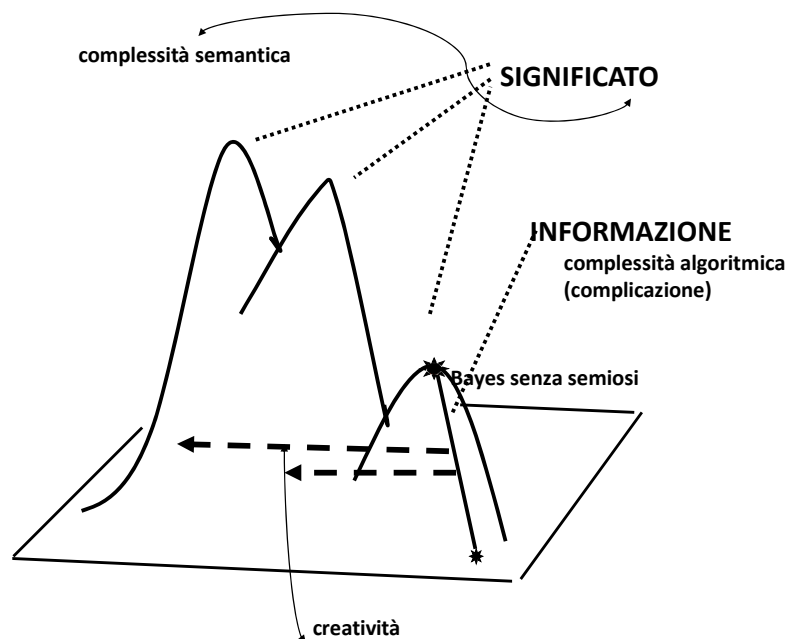


Fig.4- - La complessità semantica come molteplicità di algoritmi per affrontare la stessa situazione –

Il primo teorema di incompletezza di Gödel (1931) può essere considerato come un salto creativo in un paesaggio complesso. Il teorema stabilisce che, per ogni teoria formalmente consistente e computazionalmente numerabile, che dimostra le verità aritmetiche, si può costruire un enunciato aritmetico vero, ma non dimostrabile nella teoria. “Dimostrabile nella teoria” significa “derivabile dagli assiomi e dalle nozioni primitive della teoria, usando la logica standard del prim’ordine”.

C’è un equivalente di questo teorema nella scienza dei computer; precisamente Turing (1936) ha dimostrato che un computer universale, per un generico input, non può decidere di fermarsi (*indecidibilità dello halting problem*).

## LA NASCITA DEI LINGUAGGI E LA COSCIENZA

Il salto da un modello a un altro sotto la guida della *semiosi* è un’operazione non algoritmica peculiare di un essere vivente in interazione con l’ambiente. Ci si pone la domanda: possiamo aspettarci una evoluzione delle macchine di calcolo, fino al punto che cambino algoritmo con una procedura adattiva? La risposta è sì all’interno di uno scenario con repertorio limitato. Il cambio finora attuato è basato su una procedura *variazionale*, in base a cui il modello seguente è una versione con piccole modifiche rispetto al precedente, il quale pertanto deve essere strutturalmente stabile, cioè sopportare delle piccole varianti senza subire catastrofi. Tale è l’algoritmo genetico di Holland (Holland, 1992).

Invece l’applicazione di variazioni a un generico algoritmo di Bayes in una situazione complessa può dar luogo a instabilità, nel senso che una piccola variazione può indurre un salto discontinuo. Ciò richiede il ricorso a un algoritmo del tutto differente, violando la gradualità



postulata sopra. Un tale *salto non algoritmico* permette al matematico creativo di catturare la verità di proposizioni compatibili con gli assiomi ma non raggiungibili con il formalismo deduttivo: è questo il nucleo del teorema 1931 di Gödel.

Non si vede come una macchina possa violare il piano sul quale è stata progettata, andando oltre i graduali cambiamenti variazionali permessi dalla strategia dell'algoritmo genetico

La nascita di un linguaggio offre una soluzione al problema su esposto. Negli umani, l'apprensione può essere codificato in un linguaggio opportuno (letterario, o musicale, o plastico). Il messaggio codificato è successivamente ripreso dalla memoria e confrontato con le formulazioni linguistiche di altri apprendimenti, modificando il modello  $P(d/h)$  in modo guidato dalla presentazione successiva di ipotesi  $h$  diverse. Ciò equivale a rivisitare la stessa situazione, ma da prospettive diverse. Nella Sezione seguente descriviamo i dettagli di questa procedura linguistica. Per ora anticipiamo che la strategia non è confinata su un insieme limitato, infatti il linguaggio umano è caratterizzato dal fatto che "*fa un uso infinito di un insieme finito di risorse*" (frase attribuita a Wilhelm von Humboldt, 1836, vedi Nowak & Krakauer, 1999), dunque, questa ri-presentazione dello stesso evento cognitivo da punti di vista diversi va considerata come un passaggio non algoritmico.

Quanto detto suggerisce di definire la *coscienza* come l'esplorazione delle varie strategie  $P(d/h)$ , scegliendo quella che meglio si adattala flusso dei dati come si presentano nelle sessioni linguistiche. La coscienza non va confusa con la *consapevolezza percettiva* (*perceptual awareness*) per la quale sono stati esplorati indicatori convenienti, detti impropriamente *NCC* (*neural correlates of consciousness*) (Koch, 2004).

In effetti, la coscienza non sembra avere uno specifico *NCC*, in quanto coinvolge varie sessioni di apprensione ciascuna caratterizzata dalla propria *NCC* (Arecchi, 2010). Questo argomento verrà approfondito nella Sezione seguente.

## IL GIUDIZIO COME PROCEDURA DI BAYES INVERSA

Abbiamo considerato un scala di tempi attorno a 1 sec in cui si realizza l'*apprensione* (A). (A) risulta dalla competizione di differenti gruppi sincronizzati, e va considerato come un presente a-temporale, perché una percezione coerente implica un ri-aggiustamento delle scale temporali dei diversi canali sensori (uditivo, visivo, ecc) che singolarmente evolvono con differenti velocità (Singer, 2007). Come abbiamo visto prima, (A) richiede l'implementazione di una strategia dinamica, il controllo del caos; tuttavia (A) è deterministico solo in parte, perché in un cervello vivo l'evoluzione deterministica va rimpiazzata con una procedura di Bayes. (A) è comune a tutti gli animali superiori ed infatti viene esplorato in laboratorio su scimmie o gatti (Singer, 2007). La durata di questo evento a-temporale di sincronizzazione può essere estesa da un conveniente feedback top-down fino a 150 sec, come riportato nel caso di soggetti allenati alla meditazione senza oggetto (Lutz & al., 2004).

Una seconda scala di tempi è associata con il confronto fra l'apprensione presente e una passata richiamata dalla memoria, entrambe codificate nello stesso linguaggio. E' ragionevole arguire che il confronto richiede tre volte il tempo richiesto da una singola apprensione (A); precisamente, 1 sec è richiesto per acquistare consapevolezza dell'ultima presentazione (che chiameremo  $d$  in vista di una procedura di Bayes), un'altro intervallo di 1 sec serve per richiamare una presentazione precedente ( $h^*$ ), e 1 sec serve a rendere  $d$  e  $h^*$  co-presenti, cioè a

congiungerli in una sincronizzazione (feature binding) comune. Chiameremo *giudizio* (B) questo confronto che avviene su 3 sec.

Entro (B) si procede sfruttando una *procedura di Bayes inversa* (Arecchi, 2010). Illustriamo i dettagli della procedura. L'apprensione (A) si costruisce per Bayes diretto; l'incognita è l'ipotesi più plausibile  $h^*$ , che risulta combinando l'algoritmo top-down  $P(d|h)$  con la conoscenza della probabilità  $P(d)$  dei dati bottom-up, come segue

$$P(h^*) = P(h/d) = P(h) \times P(d/h) / P(d)$$

Al contrario, quando confrontiamo un brano  $d$  di un testo con un brano precedente  $h^*$  richiamato dalla memoria (si pensi a due versi successivi di una poesia o a due misure successive di una melodia), l'incognita è ora l'algoritmo più appropriato per armonizzare  $d$  e  $h^*$ . Esso risulta come soluzione della relazione di Bayes inversa, cioè,

$$P(d|h) = P(d) \times P(h^*)/P(h) .$$

Questa procedura, che è esclusiva degli umani perché richiede la codifica degli apprendimenti in un linguaggio simbolico, è rappresentata in Fig.5..

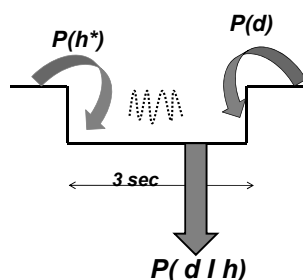


Fig 5-Giudizio come procedura di Bayes inversa. Confronto di  $d$  con  $h^*$ , da cui il modello più adeguato  $P(d|h)$  emerge a posteriori, invece di essere presupposto a priori come in Fig.4.

In tal modo, recuperiamo un punto cruciale della filosofia cognitiva. La formulazione cognitiva di Tommaso d'Aquino (1269) implicava il potere di cogliere le *cose*. Invece la formulazione di Galilei (1612) rigetta la nozione di *cosa* come priva di significato e la rimpiazza con la nozione di *oggetto* come collezione di "affezioni quantitative" cioè di aspetti misurati da apparati affidabili e pertanto validi per qualunque osservatore. Dal 1612, la scienza moderna è stata costruita come un insieme di relazioni matematiche fra i numeri che codificano le misurazioni.

Oggi, l'*oggetto* come una collezione di numeri è diventato la familiare nozione di codice a barre che classifica un prodotto in un negozio. Tuttavia non mi risulta che qualcuno provi piacere nel contemplare il codice a barre di una mela piuttosto che afferrare una mela reale. L'esplorazione della complessità ha mostrato le limitazioni della nozione di *oggetto*; la

procedura inversa di Bayes recupera la *cosa* i cui aspetti *d* sono condizionati dal punto di osservazione *h* attraverso  $P(d | h)$ .

Una scienza basata su oggetti può essere delegata a un computer, nel senso che un programma di computer può ricostruire le relazioni fra i vari aspetti quantitativi. Questa era l'affermazione di Herbert Simon, supportata dal Programma Bacon, che ricava le leggi di Keplero partendo dai dati astronomici disponibili all'inizio del XVII Sec. (Langley&al.,1987). In effetti, il problema kepleriano è del tutto privo di complessità; per contro, se affrontiamo problemi complessi, non ci aspettiamo che un computer possa rimpiazzare la creatività scientifica (Arecchi,2007 b).

Il ri-aggiustamento dei nostri codici mentali alla cosa è infatti la definizione tecnica di verità in Tommaso d'Aquino : *La verità è la conformità dell'intelletto alle cose (Veritas est adaequatio intellectus et rei)*. Fra l'altro, il ri-aggiustamento del punto di osservazione in una sequenza di Bayes inversi provvede una soluzione alla paradossale affermazione di Platone, che i nostri sensi ci ingannano perché siamo come prigionieri in una caverna, costretti a vedere solo le ombre della realtà proiettate sul fondo della caverna. In effetti, se continuiamo ad osservare le ombre sotto diverse angolazioni, confrontando le osservazioni raccolte ad ore diverse del giorno ( il confronto fra i diversi punti di vista che facciamo in (B)), finiamo con il costruire una immagine adeguata della realtà.

La ricerca sperimentale esplora con varie tecniche i correlati neurali della consapevolezza (NCC) (Koch,2004). Gli NCC possono visualizzare il reclutamento di gruppi neurali per compiti legati all'apprensione (intervalli di 1 sec). Nel caso del giudizio, abbiamo tre unità separate di apprensione (quella relativa alla codifica di *d*, quella corrispondente al richiamo di  $h^*$  e quella in cui le due apprensioni codificate sono messe a confronto per formulare un modello interpretativo (Fig.5). Come discusso da Koch (2004), dobbiamo aspettarci NCC diversi per le tre unità da 1 sec, pertanto è impossibile avere un NCC comprensivo per l'intero giudizio.

La *coscienza*, come intesa in NCC, significa *consapevolezza* di una specifica apprensione. Quando decidiamo un'azione motoria, questa consapevolezza può manifestarsi con un ritardo rispetto alla comparsa dei potenziali di azione che stimolano i muscoli (Libet,2004). Questo fatto è stato considerato come evidenza sperimentale della non esistenza di una volontà libera, in quanto diveniamo consapevoli di una decisione che si è già attuata senza aspettare il nostro consenso.

Invece della precedente definizione, definiamo la coscienza, o meglio la *auto-coscienza* come la consapevolezza di un agente di essere lo stesso giudice che sottopone a scrutinio sia l'ultimo brano *d* di discorso sia il brano richiamato  $h^*$  per costruire a posteriori una connessione  $P(d | h)$ . Questa connessione a posteriori provvede una guida per scoprire le relazioni profonde fra i brani di un testo linguistico (poesia, musica, pittura ,ecc) o di una situazione vissuta che richiede decisioni etiche da parte nostra. Queste decisioni, essendo il risultato di un giudizio, sono libere nel senso che dipendono da un impegno personale e nonerano incluse (né esplicitamente né implicitamente) nella situazione stessa.

Da queste considerazioni emerge che una decisione etica richiede un tempo ben più lungo dei tempi di Libet, e pertanto sfugge a quella inversione di ordine riportata prima.

## LA CREATIVITA'

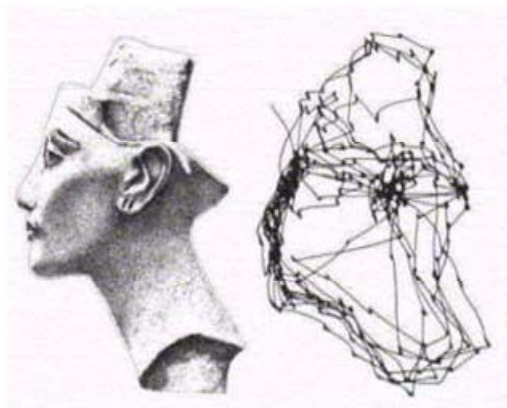
Come discusso con riferimento a Fig.5, il giudizio consiste nell'estrazione a posteriori dello stesso strumento inferenziale  $P(d | h)$  che nell'apprensione andava assegnato a priori.

Qui,  $d$  e  $h^*$  sono blocchi di dati raccolti a tempi differenti, ad esempio, due versi consecutivi di una poesia, o due misure consecutive di una melodia, o due aree distinte di un dipinto messe a fuoco in due fissazioni oculari consecutive. I due blocchi a confronto devono essere codificati nello stesso linguaggio. Inoltre devono essere adiacenti, da cui il vincolo dei 3 sec. Se allunghiamo l'intervallo cognitivo a più di 3 sec, non approfondiamo il confronto, ma piuttosto introduciamo altri blocchi di dati (Poeppel, 2004). Se dobbiamo aumentare il dettaglio di  $P(d | h)$ , dobbiamo ripetere la sessione con gli stessi  $d$  e  $h^*$ , finché si abbia un risultato soddisfacente.

Concludiamo con alcune considerazioni sulla *creatività*. Se identifichiamo la *creatività* con la scoperta di una nuova connessione  $P(d | h)$ , allora il tempo richiesto per il confronto dei due blocchi ne rappresenta un aspetto cruciale. Finora abbiamo inteso la *creatività* come la più appropriata interpretazione della relazione fra due blocchi entrambi dati. Ciò corrisponde alla lettura sensata di un testo già disponibile. Gli stessi passaggi creativi sono anche il cuore di una nuova produzione; in altre parole, un nuovo  $P(d | h)$  ha un ruolo fondamentale non solo nel costruire un ponte fra il verso  $n$ -mo e il verso  $(n+1)$ -mo di una poesia data, ma anche nell'ispirare il verso  $(n+1)$ -mo una volta che il senso profondo del verso  $n$ -mo sia stato afferrato attraverso la sua relazione col verso  $(n-1)$ -mo.

Il ponte verso un brano ancora non esistente può essere esteso a ogni produzione artistica (musica, pittura) ed è anche la forza trainante della creatività scientifica e delle decisioni etiche. Dunque, la creatività come qui considerata suggerisce una rivisitazione dell'autonomia nella scoperta scientifica e della libertà nelle decisioni etiche.

La Fig.6 dà un esempio sperimentale di come la lettura di un'opera figurativa non sia istantanea, ma richieda un tempo adeguato per effettuare il confronto fra "brani" diversi del testo. La Fig. 7 illustra la differenza fra un rimando circolare e una crescita progressiva di significati (elica) fra due brani di testo, o due persone in mutuo dialogo.



*Fig.6. Lettura di un'opera figurativa. I punti rappresentano fissazioni successive; le linee i rapidi moti saccadici (dei muscoli oculari). In questo esperimento (Noton&Stark, 1971) non venivano misurate le durate delle fissazioni.*

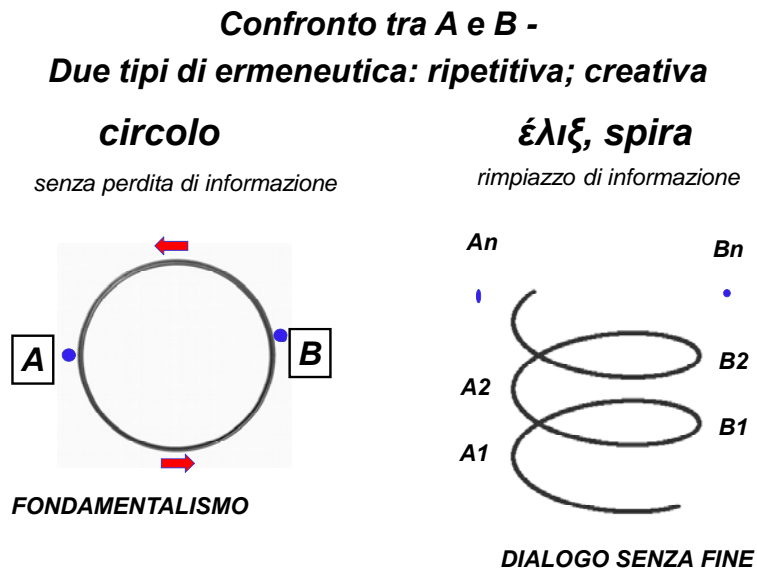


Fig. 7-Lettura di un brano di testo B in base a quanto già catturato da un brano precedente A. Nel caso a sinistra, si ha una lettura univoca di A e B, per cui i rimandi successivi si chiudono in un cerchio; nel caso a destra, il cogliere nuovi significati in B permette di recuperare nuovi significati, prima non accessibili, in A.

## BIBLIOGRAFIA

- Arecchi, F.T.(2004). Chaotic neuron dynamics, synchronization and feature binding. *Physica A* 338, 218-237 .
- Arecchi,F.T.(2007a).Complexity , Information Loss and Model Building: from neuro- to cognitive dynamics. *SPIE Noise and Fluctuation in Biological, Biophysical, and Biomedical Systems* – Paper 6602-36
- Arecchi,F.T.(2007b). Physics of cognition: complexity and creativity. *Eur.Phys.J. Special Topics* 146,205
- Arecchi, F.T.(2010). Dynamics of consciousness: complexity and creativity. *The Journal of Psychophysiology* ,24 (2),141-148
- Bayes,Th.(1763). An Essay toward solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London* 53, 370-418
- Carpenter G.A. & Grossberg, S. (2003). Adaptive Resonance Theory.  
In M.A. Arbib ( Ed), *The Handbook of Brain Theory and Neural Networks, Second Edition* (pp. 87-90). Cambridge, MA: MIT Press

- Dehaene, S. & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework, *Cognition*, 79, 1-37 (2001)
- Galilei, G. (1612). Terza lettera a M. Welser sulle macchie solari, *Opere*, vol. V (pp. 187-188). Firenze: Edizione Nazionale, Barbera 1968
- Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I, *Monatshefte für Mathematik und Physik*, 38, 173–198
- Grossberg, S. (1987). Competitive learning: from interactive activation to adaptive resonance. *Cognitive Science*, 11, 23-63;
- Holland, J.H. (1992). *Adaptation in Natural and Artificial Systems: 2nd edition*. Cambridge, MA: MIT Press.
- Koch, C. (2004). *The quest for consciousness: a neurobiological approach*. Englewood, US-CO: Roberts & Company Publishers.
- Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.
- Langley P., Simon H.A., Bradshaw G.L. & Zytkow J.M. (1987). *Scientific discovery: Computational Explorations of the Creative Processes*. Cambridge, MA: The MIT Press
- Libet, B. (2004). *Mind time: The temporal factor in consciousness*, Cambridge, MA: Harvard University Press .
- Lonergan, B. (1957). *Insight*. Toronto: University of Toronto Press.
- Lutz, A., Greischar, L.L., Rawlings, N.B., Ricard, M., & Davidson, R.J. (2004). Long-term meditators self-induce high-amplitude gamma synchrony during mental practice, *Proceedings of the National Academy of Sciences USA*, 101, 16369-16373
- Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, 171, 308–311
- Novak, M.A. & Krakauer, D.C. (1999). The evolution of language, *Proceedings of the National Academy of Sciences USA*, 96, 8028-8033
- Ott, E., Grebogi, C., & Yorke, J. (1990). Controlling chaos. *Physical Review Letters*, 64, 1196-1199.

- Pöppel, E.(1997a).A hierarchical model of temporal perception.*Trends in Cognitive Sciences*,1,56-61
- Pöppel, E.(1997b). Consciousness versus states of being conscious. *Behavioral and Brain Sciences*,20, 155-156
- Pöppel, E.(2004). Lost in time: a historical frame, elementary processing units and the 3-second window. *Acta Neurobiologiae Experimentalis*,64, 295-301
- Pöppel,E. (2009). Pre-semantically defined temporal windows for cognitive processing. *Philosophical Transactions of the Royal SocietyB*,364, 1887-1896
- Rodriguez, E., George, N., Lachaux, J-P., Martinerie, J., Renault, B.& Varela F.J. (1999) . Perception's shadow: long distance synchronization of human brain activity.*Nature*, **397**, 430- 433
- Sebeok, T. A., & Umiker-Sebeok, J. (Eds.) (1992). *Biosemiotics. The Semiotic Web*  
Berlin and New York: Mouton de Gruyter.
- Singer,W.,(2007) . Binding by synchrony, *Scholarpedia*,2(12), 1657 .
- Tommaso d' Aquino (attorno al 1269),*Summa theologica, Parte Prima*,Questione 16
- Turing, A.(1936). On computable numbers with an application to the Entscheidungsproblem *Proceedings of the London Mathematical Society*, Series 2 ,42, pp.230-265
- Wolpert,D.M., Ghahrami,Z. & Jordan,M.J. (1995). An internal model for sensorimotor Integration.*Science* 269,1880-1882.
- Womelsdorf ,T. & Fries,P. (2007) The role of neuronal synchronization in selective attention, *Current Opinion in Neurobiology*,17, 154-160

### ***Phenomenology of consciousness: creativity and liberty***

#### ***Abstract***

Today's concern for the apparently limit-less power of technologies entails an ethical defense through the so-called *principle of responsibility*. On a purely procedural level, the question arises : is technology really limit-less, or is it constrained by in-principle boundaries? As prototype of an automaton, let us consider a Universal Turing Machine; can it realize all tasks accessible to a human mind?

In this work, we explore complex problems the solution of which requires a *creativity* peculiar of humans and absent in machines.

We explore two different moments of human cognition, namely *apprehension* (A), whereby a coherent perception emerges by recruitment of large neuron groups and *judgment* (B), whereby memory retrieval of different (A) units coded in a suitable language and comparison of them leads to the formulation of a *judgment*.

The first one (A), of duration around 1 sec, appears as an a-temporal present whereby the cognitive agent acquires awareness of the environmental features and reacts consequently. A NCC (neural correlate of consciousness) can be associated with (A). (A) may be described as an interpretation of sensorial stimuli in terms of some stored algorithm, via a Bayes procedure.

The second one (B) entails the comparison of two apprehensions acquired at different times, coded in a given language, and retrieved by memory. It lasts around 3 sec and requires *self-consciousness*, as the judging agent must be well aware to be the same oneself who faces the two coded apprehensions under scrutiny in order to extract a mutual relation. At variance with (A), (B) does not presuppose an algorithm, but it rather builds a new behavioural model by an inverse Bayes procedure. It will be shown how this build up of a novel model is related to *creativity* and *free will*.

### **Key words**

Neuronal synchronization, complexity, apprehension, judgment, Bayes inference, inverse Bayes.