



UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA



**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**  
**CORSO DI LAUREA IN INGEGNERIA DELL'INFORMAZIONE**

***La Conscious Turing Machine***  
***come modello di sviluppo per l'intelligenza artificiale***

**Relatore: Prof. Gianfranco Bilardi**

**Laureanda: Sonia Pilar Gallo**

**ANNO ACCADEMICO 2022 – 2023**

**Data di laurea 13/03/2023**



*Dedicata ai miei nonni*



# INDICE

<b>INTRODUZIONE</b>	<b>3</b>
<b>MODELLO DELLA CONSCIOUS TURING MACHINE</b>	<b>5</b>
1.1 La Global Workspace Theory	5
1.2 Definizione e struttura della CTM	6
1.3 Complessità computazionale e tempi di ritardo	10
1.4 Dinamiche predittive	11
1.5 Definizioni della coscienza	12
<b>I PROCESSORI COINVOLTI NELLA COSCIENZA</b>	<b>13</b>
2.1 Linguaggio interno	13
2.2 Modello del mondo	14
2.3 Perdita della motivazione	15
<b>LA CTM NEI PROCESSI LEGATI ALLA COSCIENZA</b>	<b>17</b>
3.1 Dolore	17
3.2 Piacere	19
3.3 Vista cieca	20
3.4 Cecità da disattenzione	21
3.5 Cecità al cambiamento	22
3.6 Illusioni	23
3.7 Creazione dei sogni	24
3.8 Stati di coscienza alterati	25

**3.9 Libero arbitrio 26**

**CONCLUSIONI 29**

# INTRODUZIONE

Sin da quando è nata la tecnologia informatica ci si è domandati se un giorno questa avrebbe potuto competere con l'intelligenza umana. Oggi, con i progressi dell'intelligenza artificiale e delle reti neurali, una tale possibilità si fa sempre più concreta. Tuttavia, sono in molti a dubitare che una tale riuscita sarà mai possibile, a partire dal fatto che ancora oggi non si abbia un'idea chiara di cosa e come si generi la coscienza. Basti pensare alle domande più frequenti che filosofi, scienziati, neurologi e così tante altre figure professionali del campo si sono posti nel corso degli anni, quali:

- I computer riescono a pensare o si limitano a calcolare?
- La coscienza è una prerogativa umana?
- La coscienza dipende dal materiale che compone il cervello umano o l'hardware del computer può replicarla?

A tal proposito, il filosofo australiano David Chalmers ha voluto definire due grandi problemi riguardanti la coscienza:

- *The Easy Problem*, ossia la costruzione di una macchina in grado di simulare i sentimenti;
- *The Hard Problem*, ossia la costruzione di una macchina che effettivamente faccia esperienza dei sentimenti.

Dunque, quella che una volta era una prerogativa di filosofi e teologi si è così espansa a diverse branche della scienza, in particolar modo alla Theoretical Computer Science (TCS), volta allo studio della teoria della computazione, semantica della programmazione e analisi computazionale.

Partendo quindi da questa prospettiva si è cercato di mettere a punto un modello computazionale astratto di coscienza chiamato Conscious Turing Machine (CTM), ispirato alla Turing Machine (TM) di Alan Turing, alla Global Workspace Theory (GWT) di Bernard Baars e alla Global Neuronal Workspace Theory (GNWT) di Dohaene & Changeux.

La definizione vera e propria della CTM venne proposta nel 2021 da Lenore e Manuel Blum, due computer scientists che ritengono che la coscienza non sia prerogativa

esclusivamente umana ma proprietà di qualsiasi sistema di calcolo propriamente sviluppato. Essi, quindi, misero a punto questo modello per spiegare come una macchina potesse raggiungere una coscienza, intesa come consapevolezza dell'esperienza e di sé. Proprio per questo la caratteristica principale di questo modello è la semplicità e non può essere in alcun modo confuso come modello di funzionamento del cervello umano né come semplice Turing Machine.

Lo scopo di questa tesi sarà quindi quello di esplorare la teoria proposta dai Blum, addentrandosi in una tematica visibile e analizzabile secondo vari punti di vista ma che, in questo caso, verrà trattata essenzialmente sul piano scientifico.

Con scopo puramente esplicativo, di seguito viene proposta una scaletta di come sarà suddiviso e affrontato il lavoro di tesi.

- Capitolo 1: ispirazione del modello da quello di Baars e analisi strutturale e funzionale dal punto di vista algoritmico e computazionale;
- Capitolo 2: approfondimento dei processori coinvolti nel far emergere la sensazione di coscienza;
- Capitolo 3: spiegazione ad alto livello di come nella CTM vengono riprodotti i fenomeni tipici associati alla coscienza, in particolare il dolore, il piacere, le tre cecità, le illusioni, i sogni, gli stati di coscienza alterati e il libero arbitrio;
- Capitolo 4: conclusioni finali del modello e analisi dei punti critici.



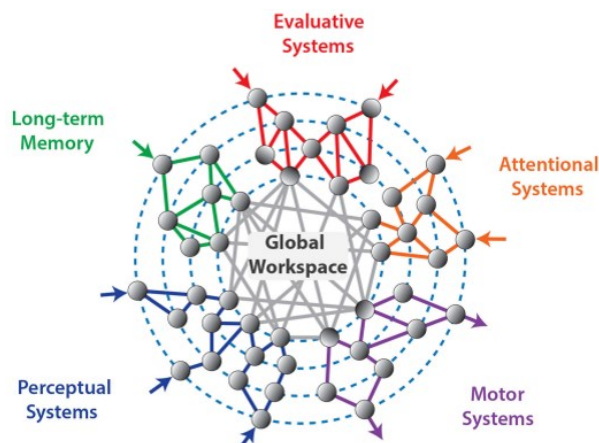
## Capitolo 1

# Modello della Conscious Turing Machine

### 1.1 La Global Workspace Theory

La Theoretical Computer Science (TCS) cercando di spiegare cosa fosse la Conscious Turing Machine (CTM) arrivò a descriverla come una semplice macchina che riprendeva in parte l'architettura cognitiva della Global Workspace Theory (GWT). Questa importante teoria venne formulata da Bernard Baars negli anni '90 per spiegare la "consapevolezza di essere coscienti" e può essere facilmente spiegata con una "metafora teatrale": in un teatro un riflettore illumina un punto sul palco dove gli attori entrano ed escono, fanno discorsi o interagiscono tra loro. Il pubblico, intento a guardare lo spettacolo, è invece al buio, così come dietro le quinte dove ci sono il regista, i macchinisti, gli sceneggiatori e simili, volti a modellare le attività del punto luminoso. Dunque, tutto ciò che è illuminato dal riflettore rappresenta i contenuti della coscienza, ciò di cui si è consapevoli, la restante parte, di contro, rappresenta i contenuti inconsci, quali i processori volti alla cognizione, alla valutazione e alla conservazione dell'informazione ma anche al controllo e alla modellazione dell'attività visibile (e dunque cosciente).

Il punto chiave di questa teoria è che Baars ritiene che la coscienza emerga nel momento in cui si connettono tra loro diverse aree del cervello, da cui "Global Workspace".



Krauss and Maier, 22 Dicembre 2020 (<https://doi.org/10.3389/fncom.2020.556544>)

La CTM riprende per buona parte il modello di Baars ma lo semplifica, considerando ad esempio la presenza di un unico attore sul palco e unificando la parte non cosciente come tutto ciò che non è illuminato, e dunque, la porzione al buio.

Nella CTM, il palco rappresenta la Short Term Memory (STM, memoria a breve termine), la quale contiene in ogni momento ciò di cui siamo consci. La restante parte rappresenta invece la Long Term Memory (LTM, memoria a lungo termine), un insieme di processori e connessioni della quale non siamo coscienti.

I processori della LTM, ognuno con le sue caratteristiche, competono tra di loro per raggiungere il “palco”, e quindi la STM, e venire così “ascoltati” dal pubblico, ovvero sia tutti gli altri processori nella LTM. La “sensazione di essere coscienti” è così spiegata dai Blum nel momento in cui l’informazione trasportata dal processore vincitore viene passata in broadcast a tutti gli altri processori. Col tempo alcuni di questi possono stabilire tra loro delle connessioni, chiamate *links*, andando così a trasformare quella che prima era una comunicazione conscia, poiché necessitava il raggiungimento della STM, a una comunicazione inconscia poiché permane all’interno della LTM.

## 1.2 Definizione e struttura della CTM

La CTM è definibile formalmente come una settina:

<STM, LTM, Up-Tree, Down-Tree, Links, Input, Output>

1. La STM è una piccola memoria in grado di trattenere un singolo “segmento d’informazione” proveniente dal processore che ha vinto la competizione.
2. La LTM è un insieme di  $N$  processori, all’incirca  $N \approx 10^7$ , inizialmente non connessi. Ciascuno di essi costituito da una memoria ad accesso casuale grande abbastanza da contenere molteplici “segmenti d’informazione”.
3. L’*Up-Tree* è un albero binario di altezza  $h$ , con  $N$  foglie, una per ciascun processore, e una singola radice per la STM. L’informazione viaggia secondo un unico percorso: parte dalle foglie e arriva verso la radice, potendo visitare al massimo  $h$  nodi.

4. Il *Down-Tree* è invece un albero di altezza 1, sempre avente la STM come radice e i processori come foglie, ma il percorso di informazione è inverso rispetto all'Up-Tree (comunicazione dalla radice alle foglie).
5. I *Links* sono comunicazioni bidirezionali che si creano e si rinforzano nel tempo quando due processori si scoprono essere utili l'un l'altro. Dal momento che i processori risiedono nella LTM anche queste connessioni vi si trovano e sono quindi responsabili della comunicazione inconscia.
6. Gli *Inputs* acquisiscono informazioni del mondo esterno attraverso alcuni sensori della CTM. Il materiale viene poi inviato a specifici processori che lo trasformano in "segmenti d'informazione".
7. Gli *Outputs*, invece, partono da essere "segmenti d'informazione" a vere e proprie istruzioni che vengono inviate agli attuatori per agire sul mondo esterno.

Finora ci siamo riferiti all'informazione messa in competizione da ciascun processore con il termine "segmento d'informazione", in realtà il termine corretto è *chunk* e si tratta di una sestina:

<address, t, gist, weight, intensity, mood>

1. *Address* è l'indirizzo del processore LTM che ha prodotto il chunk.
2. "*t*" è il momento in cui il chunk è stato prodotto e inserito nella competizione.
3. *Gist* è l'informazione vera e propria che il processore vuole comunicare tramite il chunk (può essere una risposta, l'immagine di un sogno, la descrizione del dolore...).
4. *Weight* è un numero reale che il processore dà al suo chunk in corrispondenza di quanto ritenga importante che la sua informazione raggiunga la radice (STM).
5. *Intensity* indica il livello di rilevanza del chunk e viene utilizzato durante la competizione nell'Up-Tree partendo da un valore pari a |weight|.
6. *Mood* parte con lo stesso valore del weight e anch'esso è utilizzato per la competizione nell'Up-Tree ma serve a indicare il livello di positività o negatività che l'informazione porta con sé.

La parte più importante è sicuramente il campo *gist* poiché contiene ciò che deve essere comunicato, indipendentemente dal tipo di comunicazione scelta: competizione, broadcast o links. I processori devono dunque sfruttare un linguaggio comune che gli

permetta di esprimersi e lo ritrovano nel *Brainish*, un linguaggio interno multimodale molto meglio sviluppato rispetto alle lingue che conosciamo nel mondo, quali inglese, italiano, cinese... Il Brainish è infatti ben più ricco e articolato, con una potenza di comunicazione enorme. Esso, infatti, deve essere in grado di rappresentare sensazioni, odori, sogni, immagini e tanto altro e si ritiene in parte responsabile della “sensazione di coscienza” che dichiariamo voler dimostrare nella CTM. La presenza del Brainish però non toglie che ogni processore abbia un suo linguaggio interno che include ma non è restrittivo al Brainish.

Ora che i singoli elementi sono stati definiti si può procedere con la descrizione della struttura e del funzionamento della CTM. In particolare, la CTM ha un tempo finito misurato in istanti (clock tick)  $t = 0, 1, \dots, T$ , con  $T \approx 10^{10}$ . Nasce quindi a  $t = 0$  e muore al tempo  $t = T$ . Ad ogni istante inizia una competizione, così ogni processore pone il suo chunk nella foglia corrispondente dell’Up-Tree cercando di raggiungere la radice.

Il modo in cui un chunk sale di livello nell’Up-Tree dipende da un circuito parallelo presente in ciascun nodo interno che impiega esattamente un tempo di clock per decidere il vincitore (locale): si tratta del *Coin-flip neuron*, un dispositivo che parte da una coppia di input e attraverso una funzione di probabilità ne tira fuori un vincitore. In particolare, gli inputs sono due numeri reali non negativi ( $a, b$ ), determinati dalla funzione di competizione  $f$  che rappresenta numericamente il livello di importanza di un chunk competitore, nello specifico:

$$f = f(\text{chunk}_{p,t,s}(v_s)) = f(\text{chunk}_L(v_s)) + f(\text{chunk}_R(v_s)) = \text{intensity}_{p,t,s} + c * \text{mood}_{p,t,s}$$

$$\text{con } -1 \leq c \leq 1$$

Dunque, dato un nodo interno  $v_s$  dell’Up-Tree, si definisce  $a = f(\text{chunk}_L(v_s))$  e  $b = f(\text{chunk}_R(v_s))$ , rispettivamente i valori numerici attribuiti al figlio sinistro e destro di  $v_s$  messi in competizione. La funzione di probabilità, invece, è definita nel seguente modo:

$$\text{Se } (a+b) > 0 \text{ allora} \quad p(a) = \frac{a}{a+b} \quad \wedge \quad p(b) = 1 - p(a) = 1 - \frac{a}{a+b}$$

$$\text{se } a = b = 0 \text{ allora} \quad p(a) = p(b) = 1/2$$

Dove  $p(a)$  e  $p(b)$  indicano rispettivamente la probabilità che vinca il chunk con valore numerico attribuito pari ad  $a$  o il chunk con valore numerico attribuito pari a  $b$ .

Proprio a causa di questa aleatorietà, dovuta all' Up-Tree probabilistico, la CTM prende il nome di CTM probabilistica. I valori messi in gioco nel *Coin-flip neuron* sono stabiliti dalla funzione di competizione  $f$ , quindi, una volta estratto il vincitore, vengono aggiornati i campi di quest'ultimo affinché possa procedere nella competizione. In particolare, quando un chunk sale di livello, il suo address,  $t$ ,  $gist$  e  $weight$  non vengono modificati, al contrario dei suoi campi  $intensity$  e  $mood$ .

Con lo scopo di comprendere meglio tale procedimento, supponiamo una generica competizione al nodo  $v_s$  tra i suoi due figli: sinistro, il  $chunk_{p(L)}$ , e destro, il  $chunk_{p(R)}$ . La sestina del nodo  $v_s$  andrà aggiornata in base al vincitore della competizione locale:

$$chunk_{p,t,s} = \langle address_p, t, gist_{p,t,s}, weight_{p,t,s}, intensity_{p,t,s}, mood_{p,t,s} \rangle$$

Supponiamo quindi che vinca il figlio sinistro ( $chunk_{p(L)}$ ), i campi verranno così aggiornati:

- $address_p = address_{p(L)}$
- $t = t$
- $gist_{p,t,s} = gist_{p(L),t,s-1}$
- $weight_{p,t,s} = weight_{p(L),t,s-1}$
- $intensity_{p,t,s} = (intensity_{p(L),t,s-1}) + (intensity_{p(R),t,s-1})$
- $mood_{p,t,s} = (mood_{p(L),t,s-1}) + (mood_{p(R),t,s-1})$

Per induzione, il chunk vincitore dell'intera competizione avrà come ultimi due campi quanto segue:

- $intensity_{p,t,h} = \sum_{\text{tutti i processori } p' \text{ in LTM}} (intensity_{p',t,0})$
- $mood_{p,t,h} = \sum_{\text{tutti i processori } p' \text{ in LTM}} (mood_{p',t,0})$

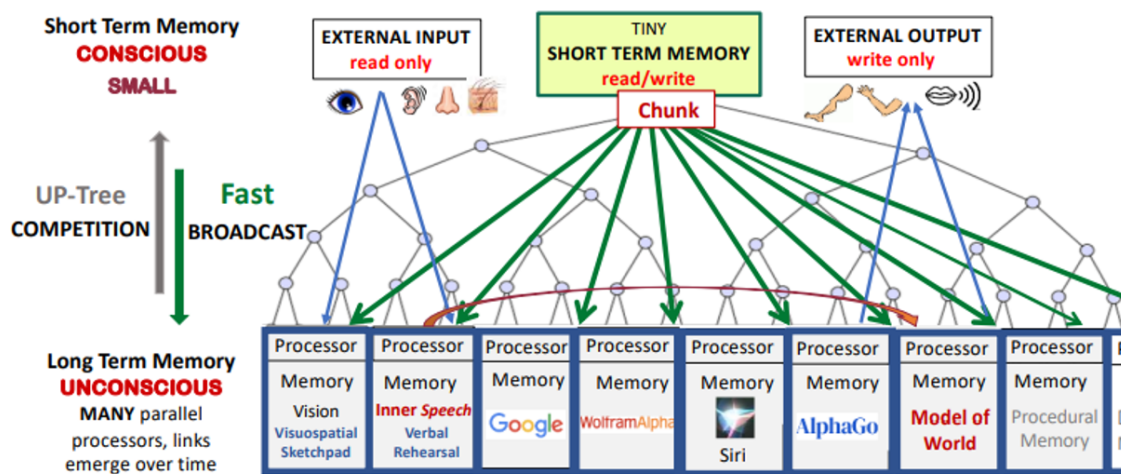
corrispondenti rispettivamente alla somma di tutte le *intensities* e *moods* dei processori alle foglie (o nodi del livello "0").

Come si può notare successivamente all'esempio fatto, la caratteristica fondamentale che definisce la funzione di competizione  $f$  è l'additività, in quanto gli ultimi due campi dei chunks vincitori sono un derivato degli stessi campi dei chunks competitori.

Dunque, i chunks messi in competizione hanno la possibilità di raggiungere la STM in proporzione al loro valore e quindi indipendentemente dalla posizione del processore che li ha generati. In particolare, la probabilità che il chunk<sub>p,t,0</sub> raggiunga la STM è:

$$f(\text{chunk}_{p,t,0}) / \sum_{\text{tutti gli } N \text{ processori } p' \text{ in LTM}} f(\text{chunk}_{p',t,0})$$

Una volta ottenuto il chunk vincitore al tempo t+h, viene sfruttato il Down-Tree per realizzare un broadcast dalla STM alla LTM, precisamente al tempo t+h+1, rendendo tutti i processori consapevoli dell'informazione trasportata dal chunk vincitore.



2022 Blum & Blum

### 1.3 Complessità computazionale e tempi di ritardo

L'aggiornamento del campo *intensity* e *mood* in un istante di tempo pone dei limiti alla sua computazione e alla sua taglia. Ad esempio, per quanto riguarda la computazione, bisogna considerare le seguenti operazioni:

- due computazioni della funzione  $f$  e il calcolo della funzione di probabilità con successiva selezione;
- copiatura dei primi quattro campi del chunk vincitore;
- somma delle *intensities* e dei *moods* e scrittura dei risultati nel chunk vincitore.

Se ci sono  $10^7$  processori e consideriamo 1 istante di tempo pari a 100ms, allora il tempo necessario affinché il chunk arrivi alla STM è pari a 2.3 secondi, poiché si tratta di un albero binario con al massimo  $h$  visite, quindi:  $h \cdot (100\text{ms}) \approx \log_2(10^7) \cdot (100\text{ms})$ .

I tempi si potrebbero ridurre drasticamente se l'albero non fosse binario ma ciascun nodo avesse più figli, ciò comporterebbe, però, un notevole discostamento dall'obiettivo dei Blum di spiegare in maniera semplice il raggiungimento di una coscienza da parte di una macchina.

#### 1.4 Dinamiche predittive

Le dinamiche predittive permettono alla CTM di regularsi sui chunks messi in competizione e sul conseguente vincitore. Il tutto avviene a livello dei processori, i quali tramite predizioni e feedback riescono ad autocorreggersi, nello specifico:

- le predizioni sono fatte da ciascun processore e riguardano i chunks trasmessi, ad esempio il valore del campo *weight*.
- i feedback provengono da chunks ricevuti e permettono il ritrovamento di errori o conferme di correttezza.
- l'apprendimento e la correzione avvengono all'interno del processore e sono necessari a un miglioramento delle predizioni.

Dunque, ciò che si verifica è un costante ciclo di predizioni, feedbacks e apprendimento all'interno della CTM. Se ad esempio un processore ritiene che il suo chunk sia importante ma nel ciclo di competizione precedente non è stato "ascoltato", può decidere di aumentarne il campo *intensity* (tramite *weight*) così da avere una probabilità più alta che il suo chunk raggiunga la radice. A tale scopo esistono una classe di algoritmi di apprendimento, gli *Sleeping Experts Algorithms* (o SEA), i quali hanno il compito di aiutare il processore a correggere eventuali errori nei suoi chunks: SEA incoraggerà il processore ad alzare il campo *intensity* del suo chunk se ciò che è stato passato in broadcast nella competizione precedente è meno importante (secondo gli SEA), mentre spingerà il processore ad abbassare il campo *intensity* in caso contrario. Tale processo avverrà solo in caso di feedback negativo, poiché se il feedback è positivo i processori non attueranno nessuna modifica.

Gli *Sleeping Experts Algorithms* svolgono quindi un ruolo fondamentale all'interno della competizione poiché sono responsabili di ciò di cui si è coscienti e ciò a cui devono prestare attenzione gli altri processori, anche considerando le comunicazioni via *links*, in

quanto il campo *intensity* è anche sintomo di quanta attenzione deve porre il processore destinatario quando legge il *gist* del chunk ricevuto.

## 1.5 Definizioni della coscienza

Per comprendere meglio cosa generi il “senso di coscienza” nella CTM è importante tenere in considerazione le seguenti definizioni formali:

- Il *contenuto cosciente* al tempo “t” è rappresentato dal chunk che vince la competizione dell’Up-Tree iniziata al tempo t-h.
- La *consapevolezza cosciente* rappresenta la ricezione in broadcast del *contenuto cosciente* da parte di tutti i processori nella LTM, considerando anche un eventuale scambio di informazioni da parte degli stessi processori tramite comunicazione inconscia (*links*).
- Il *flusso di coscienza* visto come l’insieme ordinato cronologicamente dei chunks passati in broadcast tramite Down-Tree.

La netta differenziazione tra *contenuto cosciente* e *consapevolezza cosciente* deriva dal fatto che si voglia enfatizzare il momento esatto in cui si ha la percezione del “senso di coscienza”, ovvero sia quando l’informazione del campo *gist* del chunk vincitore giunge ai processori della LTM. Dunque, il vero “senso di coscienza” non arriva al tempo t, bensì al tempo t+1.



## Capitolo 2

# I processori coinvolti nella coscienza

La sensazione di coscienza è un fenomeno interno all'individuo che si riferisce alla consapevolezza di avere esperienze interne, come ad esempio percezioni, emozioni e pensieri. Nella teoria della Conscious Turing machine, a ripresa del modello di Baars, questa sensazione emerge dalla collaborazione di più aspetti: il *Brainish*, l'architettura (intesa come struttura e funzionamento della CTM) e le dinamiche predittive.

Tuttavia, molto importante e ancora non trattato è il ruolo specifico dei processori coinvolti nella coscienza. A tal proposito, i Blum hanno stilato una lista basandosi su conoscenze empiriche nel campo della neuroscienza. Hanno quindi scoperto quali processori sono indispensabili alla coscienza e che tutti gli esseri umani hanno:

- Il linguaggio interno capace di creare e tenere traccia di quanto avviene nell'ambiente circostante realizzando piani futuri in funzione di questo.
- La conoscenza di sé stessi, fondamentale per distinguere sé stessi dal resto del mondo.
- La motivazione, un binomio di desiderio ed energia, necessario a portare a termine i compiti fondamentali alla sopravvivenza (nell'uomo corrisponderebbe al mangiare, amare, guadagnare...). Esso non è restrittivo a un singolo processore ma richiede la collaborazione di più processori.

### 2.1 Linguaggio interno

Il processore del *linguaggio interno* estrae qualunque discorso inviatogli tramite broadcast dalla STM e lo invia alle stesse locazioni raggiunte dagli inputs quando si ha una ricezione di un discorso esterno. Per questo tale processore è inteso come la capacità di parlare a sé stessi ed è fondamentale per ricordare il passato, predire il futuro e realizzare piani.

Molto simile al funzionamento di tale processore, sono i processori di *visione interna*, *udito interno* e *sensazione tattile interna*, intesi rispettivamente come la capacità di creare immagini, suoni e sensazioni tattili interne. Il fatto che le informazioni arrivino tutte alle stesse locazioni di input rende difficile capirne la provenienza, per questo il cervello ha dei trucchi per proteggere l'uomo da eventuali immagini, suoni o discorsi interni, come ad esempio far dimenticare i sogni per evitare che vengano creduti come fatti reali. I malati di schizofrenia non hanno invece alcuna abilità nel distinguere un discorso interno da uno esterno, e vale anche per immagini, suoni o sensazioni, per questo la loro patologia si manifesta con allucinazioni, deliri, comportamento disorganizzato...

## 2.2 Modello del mondo

Il *Modello-del-mondo* è inteso come una collezione di più processori volti alla costruzione dei modelli del mondo interno ed esterno della CTM. Definiamo quindi il mondo interno come un'approssimazione di come la CTM rappresenta sé stessa e il mondo esterno come una rappresentazione dell'ambiente che la circonda. A questo punto ogni cosa presente nei modelli viene etichettata e descritta tramite *gist* in Brainish insieme alle sensazioni che può avere e le azioni che può compiere. La parte più interessante è come la CTM apprende la capacità di distinguere ciò che è da ciò che non è, da qui la dicotomia tra mondo interno ed esterno.

Il processore *Modello-del-mondo* per determinare cosa la CTM è da cosa non è, sfrutta i broadcast dei chunks vincitori che sono immediatamente seguiti, tramite attuatori, da un'azione svolta nell'ambiente esterno: se uno stesso chunk passato più volte in broadcast porta ripetutamente alla stessa azione nel mondo esterno allora vuol dire che l'attuatore è parte della CTM.

Il senso di coscienza che si viene a generare nella CTM è in gran parte dovuto a questo processore. Tuttavia, non si può dire che un processore sia dotato di una coscienza o senta di essere cosciente, poiché i processori non sono altro che delle macchine che svolgono degli algoritmi. Se però si prende la CTM nella sua interezza, come un unico essere, allora esso si può dire cosciente nel momento in cui il processore *Modello-del-mondo*

rappresenta nei suoi modelli la CTM come essere cosciente e trasmette questo pensiero in broadcast a tutti gli altri processori.

Tuttavia, questo processore ha anche altri importanti compiti, come generare, ricordare e mantenere mappe personali dei mondi della CTM; predire e correggere azioni future e fare piani nel mondo esterno che esegue con la collaborazione di altri processori, tra i quali quelli citati precedentemente (*linguaggio interno, visione interna, udito interno e sensazione tattile interna...*).

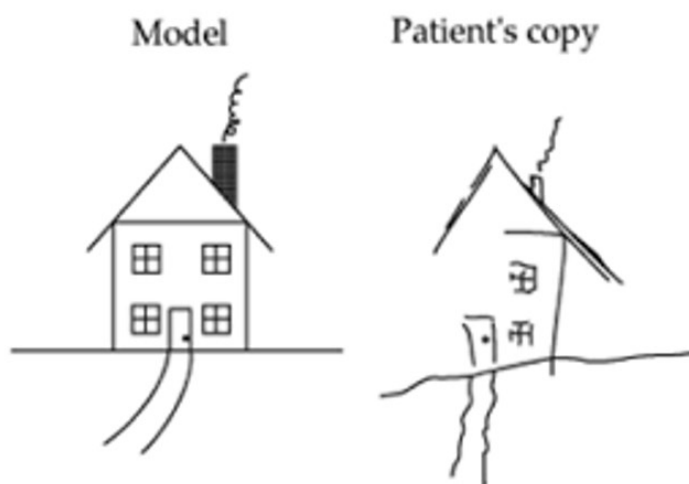
Va sottolineato però che l'esperienza del mondo esterno da parte della CTM non avviene in maniera diretta e consapevole tramite la STM ma ciò che la CTM vede, sente o percepisce nei suoi modelli del mondo avviene tramite una comunicazione diretta tra *inputs* e i processori direttamente interessati nella LTM. Questo avviene perché la conoscenza del mondo esterno non avviene in maniera consapevole, ma in maniera inconscia, esattamente come avviene per l'essere umano.

Ora che i principali processori coinvolti nel raggiungimento di una coscienza sono stati definiti e approfonditi, si ritiene necessario determinare, anche se in maniera più superficiale, quali processori sono invece stati scoperti essere del tutto non necessari allo sviluppo di una coscienza. Il più importante è sicuramente quello adibito alle emozioni, in particolare: l'imbarazzo, la paura, il senso di colpa, l'odio, la rabbia e l'amore. Anche in questo caso le scoperte si sono basate su ricerche nel campo della neuroscienza: si è visto infatti come i pazienti affetti da alessitimia, quindi privi di una consapevolezza cosciente (incapacità a identificare o descrivere le emozioni proprie e altrui), siano comunque in grado di possedere una coscienza, la quale però si sviluppa a livello di subconscio.

### 2.3 Perdita della motivazione

Alla base della sopravvivenza umana risiede la motivazione, anche se a prima vista non sembrerebbe essere necessaria allo sviluppo di una coscienza. Senza motivazione però non si riuscirebbe a individuare cosa sia necessario a sé stessi per continuare a vivere: l'uomo mangia perché sente il bisogno di mangiare. Si è visto che molte persone hanno una motivazione divisa tra quella che si prende cura della parte destra del corpo e quella

che si prende cura della parte sinistra. Un'eventuale perdita di una delle due comporta una totale perdita di coscienza della parte del corpo corrispondente. A prova di ciò è stato fatto un esperimento in cui si è chiesto a una persona con la sindrome da negligenza spaziale unilatera<sup>1</sup> di copiare un disegno, il risultato è stato il seguente:



*Articolo: Diagnosi e riabilitazione del neglect, Cosimo di Corato, Marzo 2016*

Come si nota, il paziente non considera un'intera porzione del disegno poiché non ne ha coscienza; permane però la consapevolezza e la percezione della parte destra del disegno corrispondente all'emicampo visivo cosciente.

Si è dunque dimostrato come la motivazione sia fondamentale per avere una coscienza, intesa come consapevolezza dell'esperienza e di sé.

---

<sup>1</sup> La negligenza spaziale unilatera è una sindrome neuropsicologica caratterizzata dall'incapacità da parte del paziente di percepire o prestare attenzione a oggetti, persone, rappresentazioni, collocati in un emicampo visivo (solitamente controlaterale alla lesione), e di agire in quel lato dello spazio (Robertson, 1999).

## Capitolo 3

# La CTM nei processi legati alla coscienza

Finora si è spiegata la struttura e il funzionamento interno della CTM; quindi, si è visto come emerge la sensazione di coscienza e quali processori sono in prima persona coinvolti in tale processo. Ciò che si analizzerà in questo capitolo è l'esplorazione di come la CTM potrebbe avere esperienza di una grande varietà di fenomeni associati alla coscienza, quali il dolore, il piacere, i sogni...

A tal proposito i Blum sostengono di voler trovare una soluzione ai problemi sottoposti da David Chalmers (*The Easy Problem* e *The Hard Problem*) anche se sottolineano di non avere tutte le risposte.

### 3.1 Dolore

Il dolore non è altro che una sensazione quanto più negativa quanto più è il danno subito. Esso può essere fisico o psicologico, ma in entrambi i casi si ha una forte sensazione di malessere.

Questa esperienza sensoriale ed emotiva diventa piuttosto complessa quando si cerca di inserirla in una macchina, fatta quindi di silicone e metallo. Il motivo per la quale si vorrebbe volere una macchina che prova dolore è perché solo attraverso il dolore e la paura si ottiene nella maniera più efficace l'istinto di conservazione: una macchina che conosce il dolore può proteggersi da ciò che la minaccia.

Al giorno d'oggi, i Blum dichiarano esserci solo robot con asimbolia del dolore, detta anche dissociazione del dolore, una patologia per cui chi ne è affetto prova dolore senza reazioni motorie o emotive, in altre parole, viene percepito il dolore ma non se ne soffre.

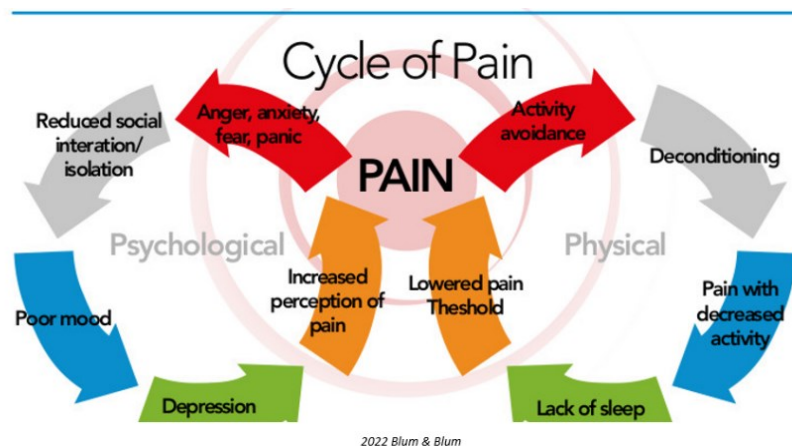
I Blum si sono quindi concentrati su cosa accade quando una persona è sottoposta a un dolore estremo: essa non è più in grado di pensare o agire perché tutta l'attenzione è rivolta sul punto da cui parte il dolore e sul dolore stesso. Da qui l'idea: quando la macchina percepisce dolore estremo, attraverso il relativo chunk che raggiunge la STM,

la sua intensità rende impossibile agli altri chunks di competere per raggiungere anch'essi la radice, almeno finché la sua intensità non è diminuita. Tutti i processori sono quindi costretti, per una frazione di tempo proporzionale all'intensità del dolore, a pensare esclusivamente a un modo per rimediare al danno.

Quando però si tratta di un dolore estremo improvviso, quale la rottura di un legamento, si ha uno shock di dolore. Ciò si traduce nella CTM come un'interruzione istantanea di tutti i processori, i quali pongono l'attività che stavano svolgendo in una *stack* e rivolgono la massima attenzione a ciò che ha causato l'interruzione. Nella STM viene proibito qualunque accesso.

La differenza tra questi due tipi di dolore sta proprio nella competizione per la STM: nel primo caso esiste ma gli altri chunk non sono in grado di accedervi, nel secondo caso, si ha un'interruzione totale delle competizioni e ci si concentra solo sul dolore. Esiste quindi una costante  $c$ , chiamata costante d'interruzione, tale per cui se l'intensità di un chunk la supera, si ottiene un blocco completo dell'Up-Tree.

Esistono poi altri due tipi di dolore: quello cronico e il cosiddetto circolo vizioso. Il primo rende difficile ma non impossibile per gli altri chunks raggiungere la STM; il secondo è un ciclo che alterna momenti in cui si prova dolore e paura. Esistono diversi cicli che possono riguardare il dolore, tutti però hanno la medesima caratteristica: il dolore sostiene e rinforza tutti gli altri elementi partecipanti. Un esempio potrebbe essere il ciclo relativo all'attacco di panico: a un disagio fisico iniziale segue l'attacco di panico vero e proprio, poi questo continua con ansia profonda e la paura che si possa avere un ulteriore attacco, così che si pongono le basi affinché il ciclo ricominci.



## 3.2 Piacere

Dolore e piacere sono spesso visti come facce opposte della stessa moneta. Proprio per questo è possibile approcciarsi al piacere nello stesso modo in cui si è fatto per il dolore. In questo caso, però, si parla di robot affetti da anedonia, ovvero sia l'incapacità di provare piacere, appagamento o soddisfazione. I Blum cercano quindi di superare questa problematica con il loro modello, in particolare, andando a studiare come reagisce un essere umano se attraversato da un momento di felicità.

Un bambino che impara che cosa è il piacere, scopre che esso può essere un antidoto contro il dolore: se sente fame allora mangia.

Un processore “esprime” piacere nello stesso modo in cui “esprime” dolore: quando un chunk di dolore arriva alla radice monopolizza la STM. Ciò è ancora più evidente nelle situazioni di piacere estremo, quando ad esempio si è sotto effetto di ecstasy.

Se un chunk del piacere raggiunge la radice, tutti gli altri processori si concentrano su un modo per aumentarne ancora di più l'intensità. Si ha dunque un aumento del piacere e una conseguente diminuzione del dolore (tramite i campi *intensity* e *mood*), inteso come un dolore generale che può essere la fame, un malessere inconscio o un dolore cronico... La CTM apprende così che le situazioni negative possono essere controbilanciate da quelle positive, e anzi, cercherà di evitare il dolore per ricercare il piacere, rendendo difatti questi due stati d'animo non simmetrici.

Tuttavia, il piacere può non essere direttamente legato al bisogno ma avere una qualità puramente edonica, ciò non apporta differenze nel comportamento della CTM quando si trova davanti uno o l'altro tipo di piacere.

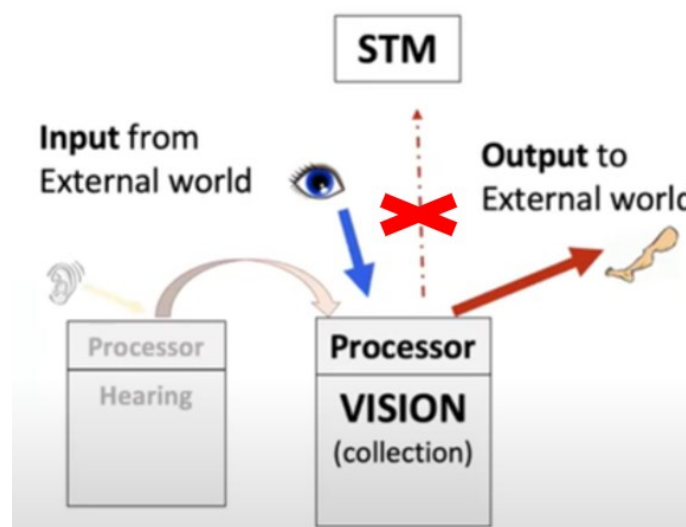
I Blum ipotizzano quindi un sistema come la CTM, costruita per mantenere il suo stato tramite le dinamiche predittive, il quale provi una sensazione di piacere nella ricerca e nel raggiungimento del suo obiettivo, ossia l'omeostasi.

### 3.3 Vista cieca

Con il termine “vista cieca” si vuole fare riferimento a quella categoria di persone cieche che riescono a muoversi, evitando gli ostacoli, senza alcun aiuto esterno.

A tal proposito è stato fatto un esperimento in cui si chiedeva a un uomo cieco di attraversare la stanza per raggiungere la porta dalla quale fosse entrato. La prima risposta fu: “ma non riesco a vedere”. Tuttavia, nel momento in cui si alzò, l’uomo riuscì tranquillamente a portare a termine la richiesta. Questo accadde poiché l’uomo vedeva ma non ne aveva consapevolezza, da qui la differenza tra coscienza e subconscio.

Dal punto di vista della CTM, gli inputs dei sensori visivi raggiungono i processori della LTM direttamente interessati ad essi. Quanto elaborato, però, non viene trasmesso alla STM e di conseguenza non viene passato in broadcast a tutti i processori. Ciò è dovuto a un generale malfunzionamento che può essere nell’Up-Tree o nel processore direttamente legato alla vista. Il risultato è il medesimo: manca il contenuto cosciente da passare in broadcast. Pur essendoci dunque l’informazione di ciò che si è visto, questa non raggiunge la STM, ma potrebbe essere che possa raggiungere altri processori attraverso la comunicazione inconscia dei *links*. Il contenuto del processore della vista è quindi inviato al processore che si occupa della camminata, il quale ha tutte le informazioni necessarie affinché possa portare a termine la richiesta evitando tutti gli ostacoli in mezzo alla stanza.



Immagini dal video, "A Framework for a Conscious AI: Viewing Consciousness through a Theoretical Computer Science Lens" (Lenore & Manuel Blum, 2022)



### 3.4 Cecità da disattenzione

La cecità da disattenzione corrisponde alla mancata capacità di notare l'esistenza di qualcosa di inaspettato quando l'attenzione è focalizzata su un qualche particolare obiettivo.

A tal proposito è stato fatto un esperimento in cui, durante la riproduzione di un video, si chiedeva di contare quante volte i giocatori con la maglietta bianca passassero la palla. Alla fine, veniva posta la domanda: hai visto il gorilla?



Immagine dal video, "The original selective attention task" (Chabris & Simons, 1999)

Questo poiché, mentre i ragazzi si passavano la palla, ad un certo punto spuntava un gorilla che attraversava la scena. Il fatto è che molte persone risposero di non averlo visto. Ciò fu dovuto al fatto che le persone erano concentrate su chi portasse la maglietta bianca, non prestando attenzione a tutto il resto, e quindi neanche al gorilla (di colore nero).



Immagine dal video, "The original selective attention task" (Chabris & Simons, 1999)

Si può quindi dedurre che neanche la CTM avrebbe potuto vedere il gorilla. Questo accade perché quando viene dato un compito alla CTM, esso raggiunge la STM e viene passato in broadcast a tutti i processori, portando a una totale modifica di tutte le intensità dei chunks competitori successivi: |weight| maggiori a ciò che serve per svolgere il compito e |weight| minori a tutto il resto.

Nel caso dell'esperimento, il processore della vista assegna massima intensità a coloro che portano la maglietta bianca e minima intensità (quasi uguale a 0) a tutto ciò che è nero. Per cui il chunk con il *gist* riferito al gorilla ha una piccolissima possibilità di entrare nella STM. Il risultato è quindi che la CTM non vede consciamente il gorilla.

L'effetto prodotto dai differenti tipi d'intensità nella CTM è consistente anche con l'abilità negli esseri umani di filtrare le informazioni irrilevanti.

### 3.5 Cecità al cambiamento

L'incapacità di notare cambiamenti, anche evidenti, è uno dei tanti fenomeni della psicologia cognitiva in grado di influenzare la *User Experience*, intesa come l'insieme di tutti gli elementi con cui un utente interagisce quando utilizza un prodotto: aspetto, colori, suoni, tatto, emozioni e tanto altro.

Per capire meglio il fenomeno psicologico si sono presi a riferimento due *frames* di un video noto perché si sono volontariamente cambiati dei dettagli nel corso della medesima scena.



Immagine iniziale e finale del video (Test Your Awareness: Whodunnit?, 2008)

A una rapida occhiata le due immagini sembrerebbero uguali. Il cervello non avrebbe il tempo di accorgersi che, in realtà, sono presenti delle importanti differenze; egli registrerebbe soltanto una scena di un crimine svolta in un salotto dove sono presenti un investigatore e dei sospettati intorno al cadavere. Questa mancata "cattura" dei dettagli è dovuta al fatto che i *frames* scorrono velocemente e dunque il cervello non ha tempo a sufficienza per registrare le minime differenze tra un *frame* e un altro.

Anche dalla prospettiva della CTM, l'informazione trasportata dal chunk descriverebbe le due immagini in maniera pressoché identica.

Questa cecità al cambiamento è dovuta al fatto che i chunks hanno una taglia massima dovuta ai limiti di tempo imposti dall'Up-Tree. Questo necessariamente si riversa nei *gists* che trasportano effettivamente l'informazione di ciò che sta accadendo. Non c'è quindi abbastanza spazio affinché ci sia una descrizione completa della scena, ciò comporta un'omissione di quanto viene ritenuto meno rilevante, in questo caso: il colore del cappotto dell'investigatore, il quadro nello sfondo ... Ciò che è presente nel *gist* è invece la descrizione della scena in generale, quindi dell'azione che sta venendo compiuta e dei protagonisti presenti in essa. Per questo le due scene sembrano pressoché identiche, in quanto i soggetti principali e l'azione compiuta rimangono uguali.

### 3.6 Illusioni

Le illusioni sono una distorsione di una percezione sensoriale o cognitiva, causata dal modo in cui il cervello è solito organizzare ed interpretare le informazioni che riceve. Esse possono coinvolgere tutti i sensi, ma quelle ottiche sono le più conosciute.

Chiari esempi di illusioni si sono visti quando si è trattato della cecità da disattenzione e della cecità al cambiamento.

Dal punto di vista della CTM si parte studiando il tipo di immagazzinamento dei *gists* ricevuti, il quale si è visto essere all'interno di memorie ad accesso casuale dei processori della LTM. Questa conservazione di informazioni permette ai processori di fare le loro predizioni tramite gli *Sleeping Experts Algorithms* e di collaborare con i processori coinvolti al fenomeno cognitivo di cui si tratterà successivamente, i sogni.

Tutti i *gists* riferiti alla vista, e appartenenti al *flusso di coscienza*, danno alla CTM la sensazione di assistere a una scena completa, quando in realtà, ciò che la CTM effettivamente vede sono solo delle immagini (piccole frazioni di una scena completa). Queste vengono collegate tra loro per trovarci un senso logico e cronologico, ricreando così la scena che però non è perfettamente aderente alla realtà e che genera quindi un'illusione, esattamente come farebbe il cervello umano.

Questo fenomeno è ottenuto grazie alla complessità dell'informazione che i *gists* riescono a trasportare e quindi grazie al *Brainish*, indispensabile per esprimere l'informazione nella maniera più completa.

### 3.7 Creazione dei sogni

I sogni sono un fenomeno psichico legato al sonno, in particolare alla fase REM (*Rapid Eye Movement*), caratterizzato dalla percezione di immagini, suoni e in generale sensazioni soggettive, le quali vengono riconosciute apparentemente reali dal soggetto sognante.

Nella CTM esiste un particolare processore volto a regolare i ritmi di sonno-veglia, così quando viene passato in broadcast o tramite *links* la necessità di riposarsi, il processore del sonno alza l'intensità dei suoi chunks per aumentare la probabilità di raggiungere la radice. Ciò si accompagna a un abbassamento dell'intensità dei chunks competitori, così, una volta raggiunta la STM, il processore del sonno ne blocca l'accesso e riduce drasticamente le intensità dei segnali direttamente collegati agli *inputs* e agli *outputs*.

Nel momento in cui il processore del sonno diminuisce le intensità dei suoi chunks, permette alla CTM di entrare nella fase REM, dunque la CTM sogna.

I sogni sono resi possibili grazie all'attivazione del processore "Creatore del Sogno", i cui *gists* raggiungono la radice trasportando con sé idee dell'attività recentemente passata della CTM insieme a immaginazioni e preoccupazioni. Quando questi chunks sono passati in broadcast, i processori rispondono ad essi come se si fossero originati nel mondo esterno, ricreando la medesima sensazione di coscienza di quando la CTM è sveglia.

L'insieme dei *gists* che si vengono ad alternare nella STM creano quindi un flusso che prende il nome di *flusso di coscienza del sonno*. Esso può essere interpretato come un "film interno" poiché costituito dall'insieme di varie esperienze sensoriali, quali vista, udito, tatto e stati d'animo misti che impattano la CTM stessa.

Ricapitolando, quindi, esiste una netta differenza tra quando la CTM dorme profondamente da quando sogna. Nel primo caso interviene il processore del sonno che

blocca qualunque accesso alla STM, nel secondo caso, invece, interviene il processore “Creatore del Sogno”, il quale permette l’accesso alla STM generando un flusso di coscienza responsabile della realizzazione di un “film interno” (il sogno).

In realtà, la presenza di un rumore forte esterno può svegliare la CTM quando questa è fase di sonno profondo, esattamente come succederebbe negli esseri umani. Dunque, esistono delle eccezioni per la quale si rende disponibile la STM a chunks appartenenti a processori diversi da quello del sonno. In generale, però, nella radice si mantiene un chunk con *gist* vuoto con lo scopo che la CTM non sia cosciente durante questa fase.

Nel momento in cui sopraggiungono i sogni, la CTM torna ad essere parzialmente cosciente. L’esperienza che ne deriva è dovuta alla potenza del *Brainish* che permette di racchiudere all’interno di singoli *gists* una quantità enorme di informazioni, proprio grazie alla completezza e complessità di tale linguaggio. Dunque, pur non avendo accesso agli *inputs* esterni, la CTM riesce a vedere, sentire e percepire eventi che non stanno accadendo realmente, ma che sembrano tanto veri quanto se la CTM li stesse effettivamente vivendo. Si è infatti già discusso nella sezione dedicata al processore del *linguaggio interno* come il cervello umano si protegga da questa somiglianza facendo dimenticare i sogni.

I collaboratori principali del processore “Creatore del Sogno” sono infatti il *linguaggio interno*, la *visione interna*, la *sensazione interna* e il processore *Modello-del-mondo*. Questi permettono dunque la realizzazione di scene e sensazioni all’interno dei mondi della CTM, testando eventualmente anche possibili situazioni di pericolo che mettono in guardia la CTM stessa su che comportamenti dovrebbe adottare e facendo dunque costanti predizioni, esattamente come se li stesse vivendo veramente.

### 3.8 Stati di coscienza alterati

Con stati di coscienza alterati si intendono quelle esperienze che l’uomo vive quando, ad esempio, sta meditando o è sotto effetto di sostanze psichedeliche. In generale si tratta di situazioni estranee da quelle di tutti i giorni, dove specialmente il modo di pensare cambia, passando da un alto senso di consapevolezza alla dissoluzione di sé.

Dal punto di vista della CTM, queste esperienze sono rese possibili grazie al processore di *Meditazione Consapevole*. Egli decide quando iniziare a meditare creando dei chunks da sottomettere all'Up-Tree, ma non è detto che riesca a raggiungere fin da subito la STM. È solo con tanta pratica e tempo che i suoi chunks aumentano di intensità e dunque aumentano le probabilità di vincere la competizione.

Una volta raggiunta la STM la difficoltà sta nel continuare a occuparla, non permettendo agli altri chunks di vincere le competizioni successive. Esattamente come avviene per gli esseri umani, anche in questo caso c'è bisogno di pratica e tempo affinché la CTM impari a rimanere concentrata e quindi a mantenere i chunks del processore *Meditazione Consapevole* all'interno della radice.

Nel momento in cui questi chunks vengono passati in broadcast, gli altri processori rispondono abbassando le intensità dei loro chunks. Anche le comunicazioni tramite *links* vengono ridotte al minimo, ciò comporta che il processore *Modello-del-Mondo* non riesca a svolgere il suo solito ruolo, ottenendo una mal comunicazione di cosa sia e cosa non sia la CTM, da cui il pensiero di dissoluzione di sé come pensiero emergente e ricorrente durante la meditazione.

### 3.9 Libero arbitrio

Il tema del libero arbitrio è da sempre al centro dell'attenzione di filosofi e teologi e indica la capacità di scegliere, pensare e agire liberamente. Col tempo si è espanso a diverse branche della scienza, in particolare alla fisica, volendo scoprire se effettivamente esistesse la possibilità di agire e pensare liberamente andando oltre le leggi fondamentali che regolano la natura.

Secondo la prospettiva della TCS (Theoretical Computer Science), che si è detto essere il filtro con la quale comprendere la coscienza, non esiste il libero arbitrio poiché nessuna decisione, per quanto libera, potrà mai violare le leggi della fisica ed essere dunque priva di qualunque vincolo.

Tuttavia, la TCS afferma che può esistere la sensazione di libero arbitrio, intesa come la consapevolezza che di fronte a una scelta si possano fare i dovuti calcoli e alla fine si

possa prendere una decisione, la quale è considerata la migliore possibile in funzione di qualcosa che può essere un obiettivo o una necessità.

Questa seconda definizione porta con sé un ampio utilizzo delle dinamiche predittive e delle limitazioni delle risorse. Nel primo caso si fa riferimento al fatto che ogni azione comporta delle conseguenze che vengono tenute in considerazione proprio per raggiungere il fine ultimo. Nel secondo caso, invece, si ragiona in termini di tempo, spazio, mole d'informazioni utili che si guadagnano ... e anche in questo caso si tratta di considerazioni fondamentali al fine di prendere la miglior decisione possibile.

Per comprendere meglio quanto appena affermato si supponga di dover sostenere un esame di matematica. L'obiettivo ultimo è quello di superare l'esame, tuttavia, esiste un grande intervallo di voti che porti a termine l'obiettivo: può infatti esistere un Marco che punti a superare l'esame con il massimo di voti o un Matteo che spera solamente di passarlo. Ciò comporterà un piano organizzativo differente: nel primo caso verrà sfruttato molto più tempo ed energia rispetto al secondo caso. Marco infatti sa che se uscirà tutte le sere prima dell'esame avrà meno probabilità di raggiungere il suo obiettivo. Dunque, per poter organizzare un piano di studio Marco e Matteo faranno delle predizioni e, parallelamente, calcoleranno un *trade off* tra le energie e il tempo che gli servirà spendere.

Questi calcoli fatti di predizioni e limitazioni di risorse vengono svolti nella CTM in maniera perlopiù inconscia. A tal proposito, si prenda come esempio una partita di scacchi in cui tocca alla CTM muovere il pezzo successivo. La necessità di sapere cosa sia meglio muovere raggiungere la STM e il messaggio viene passato in broadcast. I processori interessati suggeriscono la mossa che ritengono più opportuna, ma solo uno è chunk il vincitore, per cui la CTM avrà conoscenza di solo alcuni suggerimenti, quelli che hanno raggiunto la radice, quando in realtà si sa che alla base dell'Up-Tree diversi processori hanno contribuito per dare la risposta finale.

Questo tipo di funzionamento fa sentire la CTM come se la sua decisione finale arrivasse da dentro di sé, senza però avere una chiara consapevolezza di come si sia originata. È proprio questo insieme di suggerimenti di risposte consce e inconsce che, secondo i Blum, fa emergere la sensazione di libero arbitrio.

Dunque, la CTM non si rende conto di avere dei vincoli poiché le risposte ultime che produce hanno una natura per buona parte inconscia. Per questo si sarebbe potuto ottenere il medesimo risultato se al posto di un Up-Tree probabilistico si fosse avuto un Up-Tree deterministico, in quanto ciò che più conta è la presenza del binomio conscio-inconscio.

L'ultimo punto che resta da discutere è il tempo di ritardo da quando il chunk viene sottoposto all'Up-Tree a quando raggiunge la radice. Esso è necessario affinché ci sia la selezione di tutti i chunks vincitori locali e si può facilmente interpretare come il tempo da quando si prende una decisione inconsciamente e la si attua consciamente.



## Capitolo 4

# Conclusioni

Il lavoro di tesi si è focalizzato su un particolare modello computazionale astratto di coscienza, la Conscious Turing Machine (CTM).

Il punto di partenza è stato lo studio dell'architettura portante del modello che ha preso ispirazione dalla Global Workspace Theory di Bernard Baars, il quale sostiene come l'emergere della coscienza dipenda dalla collaborazione di più aree del cervello. Nel caso del modello è dunque l'insieme di tutti i componenti che lo costituiscono a produrre tale sensazione.

Entrando nello specifico, si è visto come la CTM sia divisa in una porzione conscia (STM, *Short Term Memory*), in grado di memorizzare una quantità limitata di informazioni, e una porzione inconscia (LTM, *Long Term Memory*) dove risiedono tutti i processori e le connessioni (*links*) che si generano tra loro.

La collaborazione tra queste due porzioni avviene tramite una competizione tra processori, i quali puntano a raggiungere la STM per poter essere “ascoltati” a livello globale da tutti i processori del modello ed eventualmente avere anche un'interfaccia verso il mondo esterno. È questo il momento esatto in cui viene generata la “sensazione di essere coscienti”, ossia quando l'informazione trasportata dal processore vincitore è passata in broadcast a tutti gli altri processori.

Questo tipo di architettura ha però reso necessario un tipo di comunicazione che fosse comprensibile e universale a tutti i processori, da qui la teorizzazione del *Brainish*, un linguaggio multimodale in grado di descrivere stati d'animo, immagini, sensazioni ... a livelli ben più complessi delle normali lingue del mondo.

Dunque, è attraverso l'architettura e il *Brainish* che la CTM ha la possibilità di interfacciarsi col mondo circostante e di regolarsi nel suo “mondo interno”, ma il come lo fa è determinato dalle dinamiche predittive, un insieme di cicli costanti di predizioni, feedbacks e correzioni svolte all'interno dei processori.

Il lavoro di tesi si è poi focalizzato su quali processori siano effettivamente coinvolti nel processo di coscienza: attraverso lo studio empirico nel campo delle neuroscienze, si sono evidenziati tre aspetti fondamentali per i quali un essere umano non avrebbe una coscienza. Questi sono stati attribuiti a tre processori diversi: uno per il *linguaggio interno*, responsabile del continuo dialogo interiore che l'uomo utilizza per pensare, agire e fare piani futuri; uno per il *Modello-del-Mondo*, volto all'identificazione di sé e uno per il processore della *Motivazione*, un binomio di energia e desiderio, essenziale per la presa di consapevolezza dell'ambiente circostante.

Il lavoro si è concluso con una interpretazione ad alto livello di come i principali fenomeni umani legati alla coscienza possano essere riprodotti dalla CTM, quali ad esempio: il dolore, il piacere e i sogni.

Dopo aver ripercorso i punti chiave del lavoro di tesi, si vuole dare uno sguardo al futuro, in particolare, ad alcuni temi che necessitano di un approfondimento. A tal proposito, viene suggerita la seguente lista:

- un'accurata e dettagliata definizione del *Brainish* che per ora è stato solo teorizzato su carta. Nel prossimo futuro si spera di trovare un modo per concretizzarlo sia a livello algoritmico che a livello informatico.
- la formulazione delle prime predizioni e il modo in cui i feedback "pesano" sulle autocorrezioni dei processori. A tal proposito, si è teorizzato degli *Sleeping Experts Algorithms* che aiutano i processori a migliorare le loro predizioni, ma non si è ancora trovata una loro implementazione pratica.
- la necessità o meno di una unità di controllo, in quanto la CTM, per quanto sia un modello semplificato di coscienza, è costituita da una grande quantità di componenti che devono essere gestiti. A tal proposito, si prenda come esempio la costante d'interruzione  $c$  (capitolo 3.1) che è in grado di instaurare uno stato di fermo di tutti i processori: questa reazione è insita nelle singole memorie ad accesso casuale di ciascun processore o il comando di stop viene dato da una cellula di controllo esterna?

È chiaro, dunque, come ci si trovi ancora agli albori nel campo dei modelli di coscienza per le macchine. Questo è dovuto al fatto che, ancora oggi, le informazioni sulla coscienza

sono ben poche sia dal punto di vista psicologico-filosofico che dal punto di vista matematico-computazionale.

Per ora quindi si mettono le basi per quello che potrà essere un futuro alternativo e completamente rivoluzionario. La domanda che resta da porsi è: si sta andando nella direzione giusta? Ai posteri l'arduo compito di rispondere.

# Bibliografia

- [1] Lenore Blum, Manuel Blum. A Theory of Consciousness from a Theoretical Computer Science Perspective: Insights from the Conscious Turing Machine. <https://doi.org/10.48550/arXiv.2107.13704>. 29 Giugno 2021 / Pag. 1-64.
  
- [2] Lenore Blum, Manuel Blum. A Theoretical Computer Science Perspective on Consciousness. World Scientific: Journal of Artificial Intelligence and Consciousness. 15 Marzo 2021 / Vol. 8 / N° 1 / Pag. 1-42.  
<https://www.worldscientific.com/doi/abs/10.1142/S2705078521500028>.
  
- [3] Lenore Blum, Manuel Blum. Insight from the Conscious Turing Machine. 2021.  
<https://www.youtube.com/watch?v=Q0wzo8eXK30&list=WL&index=45>.
  
- [4] Lenore Blum, Manuel Blum. A framework for a Conscious AI. 2022.  
<https://www.youtube.com/watch?v=q7A2M4dauzk&list=WL&index=46>.
  
- [5] Lenore Blum, Manuel Blum. Alan Turing and the other theory of computing and can a machine be conscious? 2017.  
<https://www.youtube.com/watch?v=5-vWptZ6kNo&list=WL&index=47>.
  
- [6] Lenore Blum, Manuel Blum. A theoretical computer science perspective on consciousness. 2021.  
<https://www.youtube.com/watch?v=kuH2GDj734A&list=WL&index=48>.
  
- [7] Lenore Blum. A Theoretical Computer Science Perspective on Consciousness by Prof Lenore Blum. 2021.  
<https://www.youtube.com/watch?v=v9-wcx3lt1c&t=1429s>.
  
- [8] Manuel Blum. Towards a Conscious in AI. 2020.  
<https://www.youtube.com/watch?v=6Xx39Efe2Yc&list=WL&index=52>.

- [9] Lenore Blum. What Can Theoretical Computer Science Contribute to the Discussion of Consciousness? 2020.  
<https://www.youtube.com/watch?v=Ci5AAsuF-HI&list=WL&index=53>.
- [10] Giorgio Buttazzo. Artificial Consciousness: Utopia or Real Possibility. IEEE. Giugno 2021 / Vol. 34 / Pag. 24-30.  
<http://retis.sssup.it/~giorgio/paps/2001/ieeecm01.pdf>.
- [11] Patrick Krauss, Andreas Maier. Will we ever have conscious machines? *Frontiers in Computational Neuroscience*. 2020 / Vol. 14 / Pag. 1-14.  
<https://doi.org/10.3389/fncom.2020.556544>.