Towards a Conscious AI
A Computer Architecture Inspired by Cognitive Neuroscience

by

Three Blums: Manuel, Lenore, Avrim

October 2018

School of Computer Science Carnegie Mellon University

What is Consciousness?

Roughly speaking, **consciousness** or **conscious awareness** is everything that you pay attention to.

Consciousness is what you are aware of

- Your senses: what you see, hear, smell, taste, touch….
- Your inner speech.
- Your dreams (but not dreamless sleep).
- Your feelings: joys, fears, sorrows, pains….

What's Conscious and what's not?

Conscious

1. See someone at a party. Know his/her name, but what is it?
2. Dreams
3. Selective attention
4. **Problem clarification**
5. To **learn to play** ping pong, you must pay conscious attention to your game.

Unconscious

1. Search for name… name can pop to (conscious) mind ½ hour later.
2. Sleep without dreams
3. Seen but not attended to
4. **Problem incubation**
5. To play ping pong in a **tournament**, best to go on automatic.

Hadamard. Poincaré. Gauss.

" One experience **Hadamard** describes: "***On being very abruptly awakened by an external noise, <u>a solution long searched for appeared to me at once without the slightest instant of reflection on my part</u> … and in a quite different direction from any of those which I had previously tried to follow***".

Another example from **Poincaré** who, putting his mathematics aside, was traveling on a geological excursion when suddenly, **as he was about to board a bus, "*<u>the idea came to me, without anything in my former thoughts seeming to have paved the way for it,</u> that the transformations I had used to define the Fuchsian functions were identical with those of non-Euclidean geometry*"**.

Another example from **Gauss** who, **<u>after years of failing to prove a particular mathematical theorem</u>, <u>finally succeeded "*by the grace of God. Like a sudden flash of lightning … I myself cannot say what was the conducting thread*</u> *which connected what I previously new with what made my success possible*".**

What is Consciousness?

This talk will present a formal model of Turing machine. We call it the
Conscious TM (CTM) or **Conscious AI (CAI)**
depending on what we want to emphasize.

What is Consciousness?

This talk will present a formal model of Turing machine. We call it the **Conscious TM (CTM)** or **Conscious AI (CAI)** depending on what we want to emphasize.

After formalizing the **Conscious TM (CTM)**, we define **consciousness** in that model, then point out properties of that "**consciousness** in the model".

What is Consciousness?

This talk will present a formal model of Turing machine. We call it the
Conscious TM (CTM) or **Conscious AI (CAI)**
depending on what we want to emphasize.

After formalizing the **Conscious TM (CTM)**, we define **consciousness** in that model, then point out properties of that "**consciousness** in the model".

The quality of a formalization and definition depends largely on how closely the notion squares with what you think it should be, and whether or not it helps you to understand the concept. You be the judge.

Our model is inspired by the works of Cognitive Neuroscientists, Psychologists and Philosophers

John Anderson, Bernard Baars*, David Chalmers, Francis Crick & **Christof Koch**, Antonio Damasio, Daniel Dennett, **Stanislas Dehaene,** Gerald Edelman, Michael Gazzaniga, Pennti Haikonen, Ryota Kanai, Stephen Laberge, Rodolfo Llinás & Urs Ribary, Drew McDermott, **Kevin O'Regan, Bjorn Merker, Allan Newell,** Don Norman & Tim Shallice, Michael Posner, Vilayanur Ramachandran, Giulio Tononi, **…**

We particularly recommend:** ***Fundamentals of Cognitive Neuroscience, **by Baars and Gage (2013).**

The Easy and Hard Problems (David Chalmers)

The **Easy Problem**: make a machine that **simulates** feelings of **pain** and **joy**.

The **Hard Problem**: make a machine that truly **experiences** feelings of **pain** and **joy**.

*Qualia = <u>Individual instances of subjective, conscious experience</u>

The Easy and Hard Problems (David Chalmers)

The **Easy Problem**: make a machine that **simulates** feelings of **pain** and **joy**.

The **Hard Problem**: make a machine that truly **experiences** feelings of **pain** and **joy**.

Chalmers' definition is a lot more general than this. He's interested in all **qualia**.*

*Qualia = Individual instances of subjective, conscious experience

The Easy and Hard Problems (David Chalmers)

The **Easy Problem**: make a machine that **simulates** feelings of **pain** and **joy**.

The **Hard Problem**: make a machine that truly **experiences** feelings of **pain** and **joy**.

Chalmers' definition is a lot more general than this. He's interested in all **qualia**.*

Our research and this talk is restricted to understanding pain and joy, including the extremes of agony and ecstasy.

*Qualia = <u>Individual instances of subjective, conscious experience</u>

The Easy and Hard Problems restricted to Pain & Joy

The **Easy Problem**: make a machine that **simulates** feelings of **pain** and **joy**.

The **Hard Problem**: make a machine that truly **experiences** feelings of **pain** and **joy**.

Chalmers' definition is a lot more general than this. He's interested in all **qualia**.

Our research and this talk is restricted to understanding pain and joy, including the extremes of agony and ecstasy.

We have a reasonably good answer for pain.

We have only a partial answer for joy.

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

We look not for complexity but for simplicity :
We are not looking for a complex model of the brain.
We are looking for a simple model of consciousness.

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

We look not for complexity but for simplicity :
We are not looking for a complex model of the brain.
We are looking for a simple model of consciousness.

- Evolution abhors parsimony, says John Anderson. Mathematics thrives on it, say we.

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

We look not for complexity but for simplicity :
We are not looking for a complex model of the brain.
We are looking for a simple model of consciousness.

- Evolution abhors parsimony, says John Anderson.
  Mathematics thrives on it, say we.
- **Our aim is to propose a simple model that we can understand and prove theorems about.**

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

We look not for complexity but for simplicity :
We are not looking for a complex model of the brain.
We are looking for a simple model of consciousness.

- Evolution abhors parsimony, says John Anderson. Mathematics thrives on it, say we.
- **Our aim is to propose a simple model that we can understand and prove theorems about.**
- **We want properties of consciousness to be emergent, not programmed in.**

The Easy and Hard Problems

First, what's the difference between **Simulation** and **Experience?**

The Easy and Hard Problems

First, what's the difference between **Simulation** and **Experience?**

The disorder called **Pain Asymbolia.**

The Easy and Hard Problems

First, what's the difference between **Simulation** and **Experience?**

The disorder called **Pain Asymbolia.**

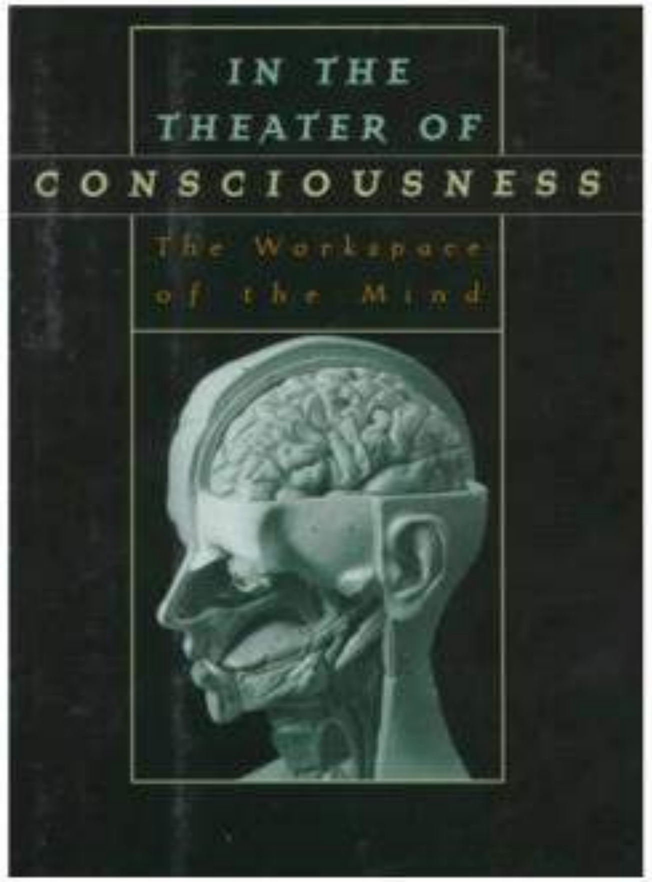
Pain Asymbolia 1. **Pain Asymbolia 2.**

Consciousness has to do with the Architecture of the Brain

The architecture presented here deals with the brain at a very high level of abstraction – a level way above that of neurons.

The architecture is NOT Obvious
It is called the Theater Model or
Global Workspace Model (GWM)

GWM is an extraordinary idea for explaining consciousness.
It is due to neuroscientist Bernard Baars



Baars' Theater of Consciousness

Bernard Baars describes conscious awareness through a **theater analogy**: consciousness is the activity of actors in a play performing on the stage of working or **Short Term Memory** (**STM**). The performance is observed by a huge audience of processors in **Long Term Memory** (**LTM**) that are sitting in the unconscious darkness.

STM = Short Term Memory **LTM** = Long Term Memory

The conscious self is not privy to the workings of the unconscious self:

The conscious self is not privy to the workings of the unconscious self:

The conscious self is not privy to the workings of the unconscious self:

What's her name?

The conscious self is not privy to the workings of the unconscious:

You recall where you first met ↗. It gets broadcast ↘.
Something about what she does ↗. It gets broadcast ↘.
Her name begins with T ↗. It gets broadcast ↘.

The conscious self is not privy to the workings of the unconscious:

You recall where you first met ↗. It gets broadcast ↘.
Something about what she does ↗. It gets broadcast ↘.
Her name begins with T ↗. It gets broadcast ↘.

A half hour later her name comes up ↗ from audience (unconscious) - which has been thinking, searching - to stage (conscious).

The conscious self is not privy to the workings of the unconscious:

You recall where you first met ↗. It gets broadcast ↘.
Something about what she does ↗. It gets broadcast ↘.
Her name begins with T ↗. It gets broadcast ↘.

A half hour later her name comes up ↗ from audience (unconscious) - which has been thinking, searching - to stage (conscious).

The conscious self - on stage - doesn't know how or where her name was found.

The conscious self is not privy to the workings of the unconscious:

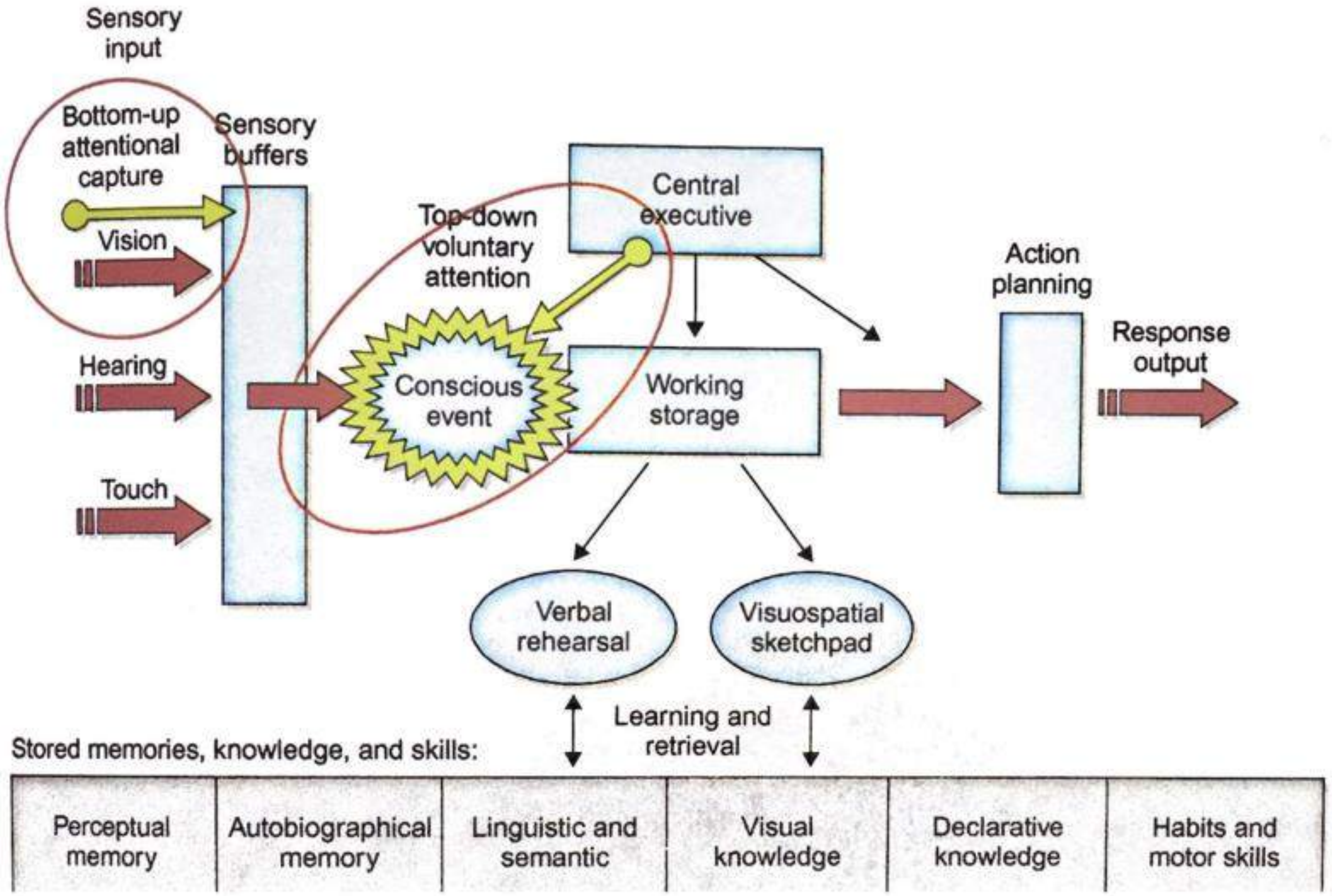
You recall where you first met ↗. It gets broadcast ↘.
Something about what she does ↗. It gets broadcast ↘.
Her name begins with T ↗. It gets broadcast ↘.

A half hour later her name comes up ↗ from audience (unconscious) - which has been thinking, searching - to stage (conscious).

The conscious self - on stage - doesn't know how or where her name was found.

Next up: **Baars' model**

Bernard Baars' Model of Consciousness



What can Theoretical Computer Science contribute to the Discussion of Consciousness?

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

- A well-defined formal model

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

- A well-defined formal model
- A good formal definition of Consciousness

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

- A well-defined formal model
- A good formal definition of Consciousness
- Explanations how Agony and Ecstasy might arise in a machine.

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

- A well-defined formal model
- A good formal definition of Consciousness
- Explanations how Agony and Ecstasy might arise in a machine.
- **Understanding** for distinguishing **simulation** from **experience.**

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

- A well-defined formal model
- A good formal definition of Consciousness
- Explanations how Agony and Ecstasy might arise in a machine.
- **Understanding** for distinguishing **simulation** from **experience.**

Without understanding, there is no way to tell if an entity (animal or robot) is conscious.

What can Theoretical Computer Science contribute to the Discussion of Consciousness?

- A well-defined formal model
- A good formal definition of Consciousness
- Explanations how Agony and Ecstasy might arise in a machine.
- **Understanding** for distinguishing **simulation** from **experience.**

Without understanding, there is no way to tell if an entity (animal or robot) is conscious.

includes humans

Next up is the formal definition of the Conscious Turing Machine (or Conscious AI)

The purpose of the Conscious Turing Machine is **NOT** to compute uncomputable functions. That is not possible.

Next up is the formal definition of the Conscious Turing Machine (or Conscious AI)

The purpose of the Conscious Turing Machine is **NOT** to compute uncomputable functions. That is not possible.

Nor is its purpose to compute functions more efficiently.

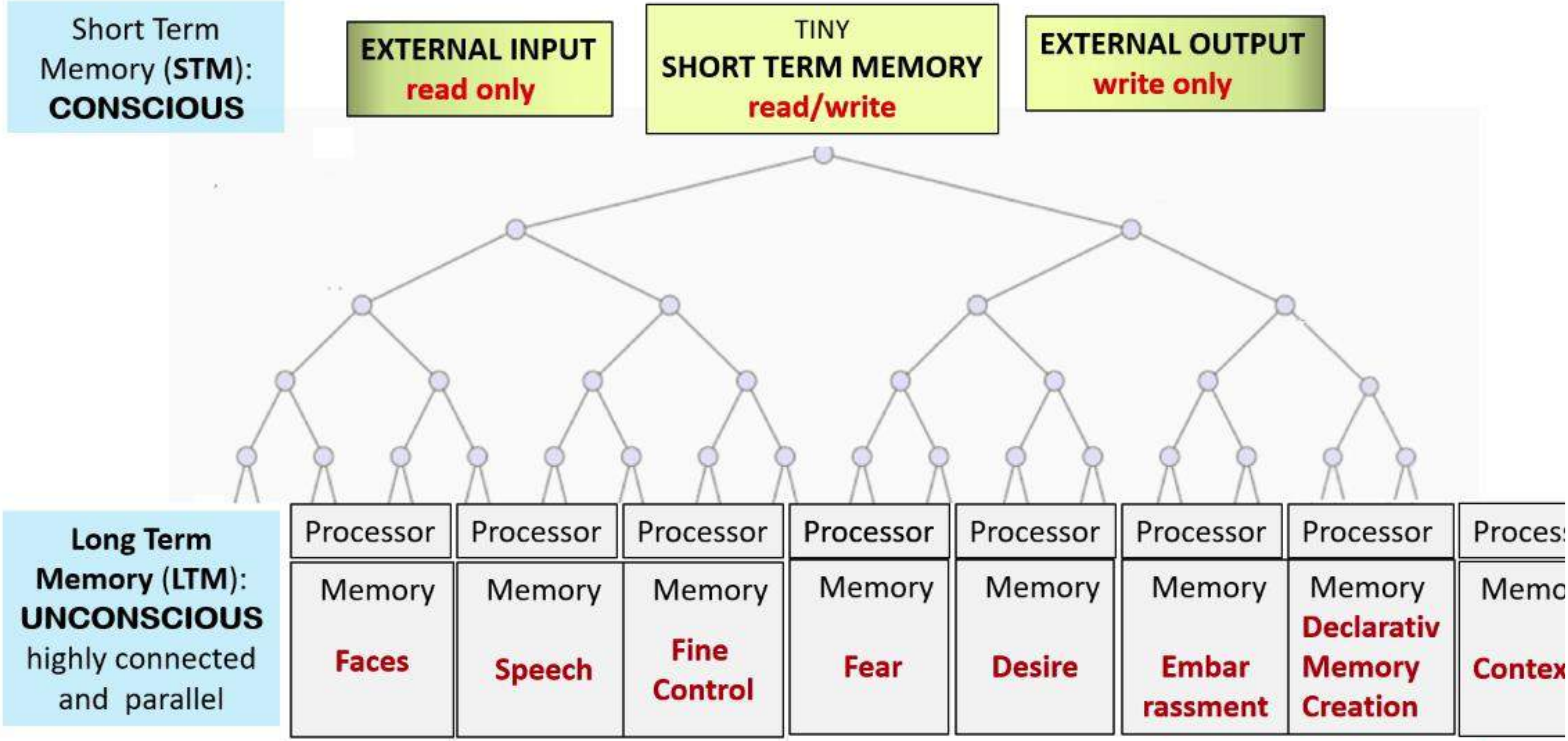
Next up is the formal definition of the Conscious Turing Machine (or Conscious AI)

The purpose of the Conscious Turing Machine is **NOT** to compute uncomputable functions. That is not possible.

Nor is its purpose to compute functions more efficiently.

Its purpose is to suggest possible solutions to the hard problem.

THE CONSCIOUS TM DEFINITION



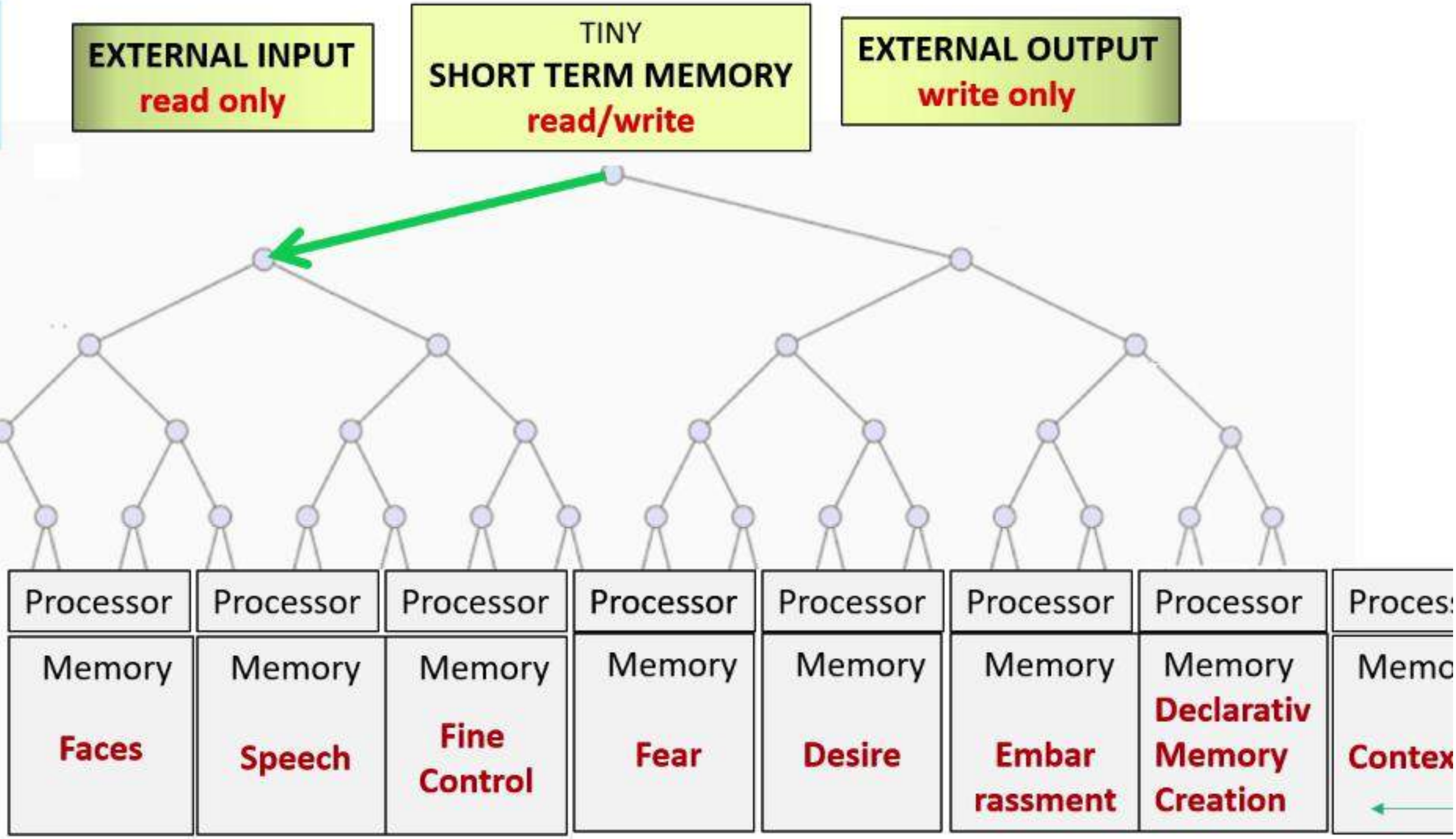
No Central Executive

THE CONSCIOUS TM DEFINITION

Short Term Memory (**STM**): **CONSCIOUS**

Central Executive function is performed by Unconscious Processors

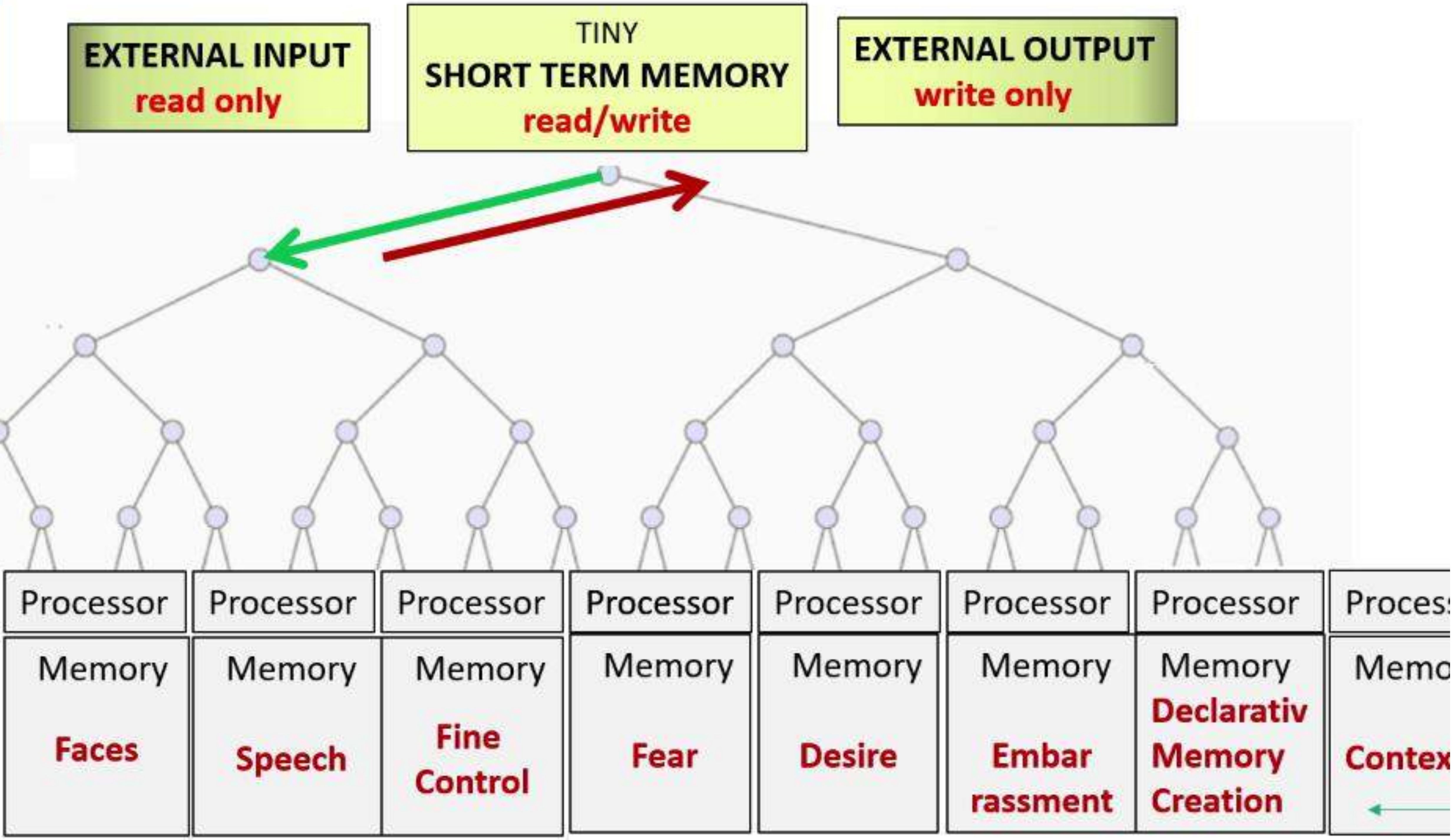
Long Term Memory (**LTM**): **UNCONSCIOUS** highly connected and parallel



THE CONSCIOUS TM DEFINITION

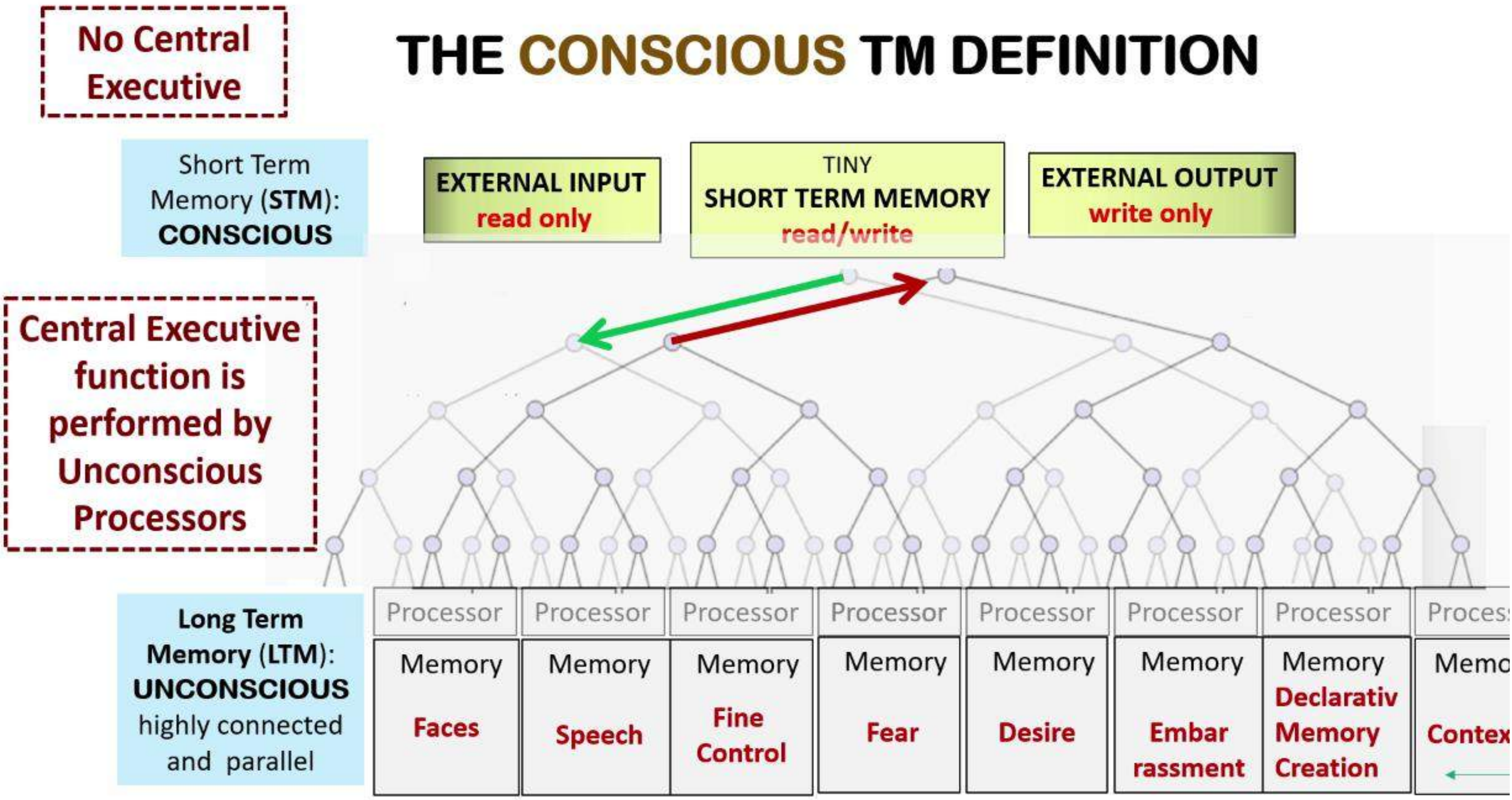
No Central Executive

Short Term Memory (**STM**): **CONSCIOUS**



Central Executive function is performed by Unconscious Processors

Long Term Memory (**LTM**): **UNCONSCIOUS** highly connected and parallel


No Central Executive
THE CONSCIOUS TM DEFINITION
Short Term Memory (STM): CONSCIOUS
EXTERNAL INPUT
read only
TINY
SHORT TERM MEMORY
read/write
EXTERNAL OUTPUT
write only
Central Executive function is performed by Unconscious Processors
Long Term Memory (LTM): UNCONSCIOUS highly connected and parallel
Processor
Processor
Processor
Processor
Processor
Processor
Processor
Memory
Faces
Memory
Speech
Memory
Fine Control
Memory
Fear
Memory
Desire
Memory
Embar rassment
Memory
Declarativ Memory Creation

No Central Executive

THE CONSCIOUS TM DEFINITION

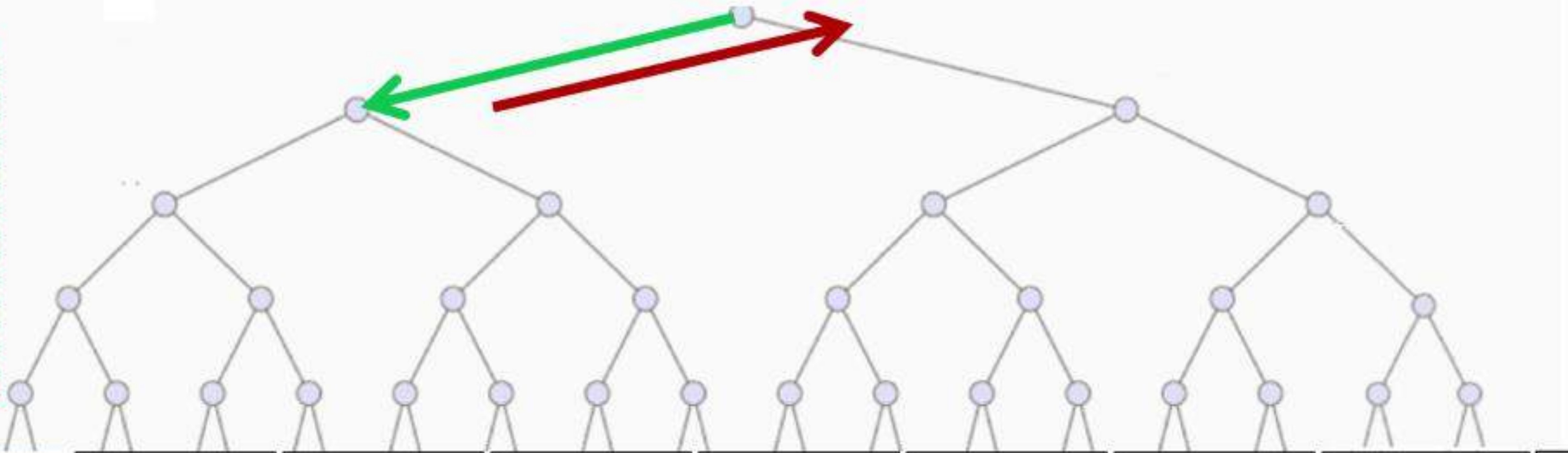
Short Term Memory (**STM**): **CONSCIOUS**

EXTERNAL INPUT
read only

TINY
SHORT TERM MEMORY
read/write

EXTERNAL OUTPUT
write only

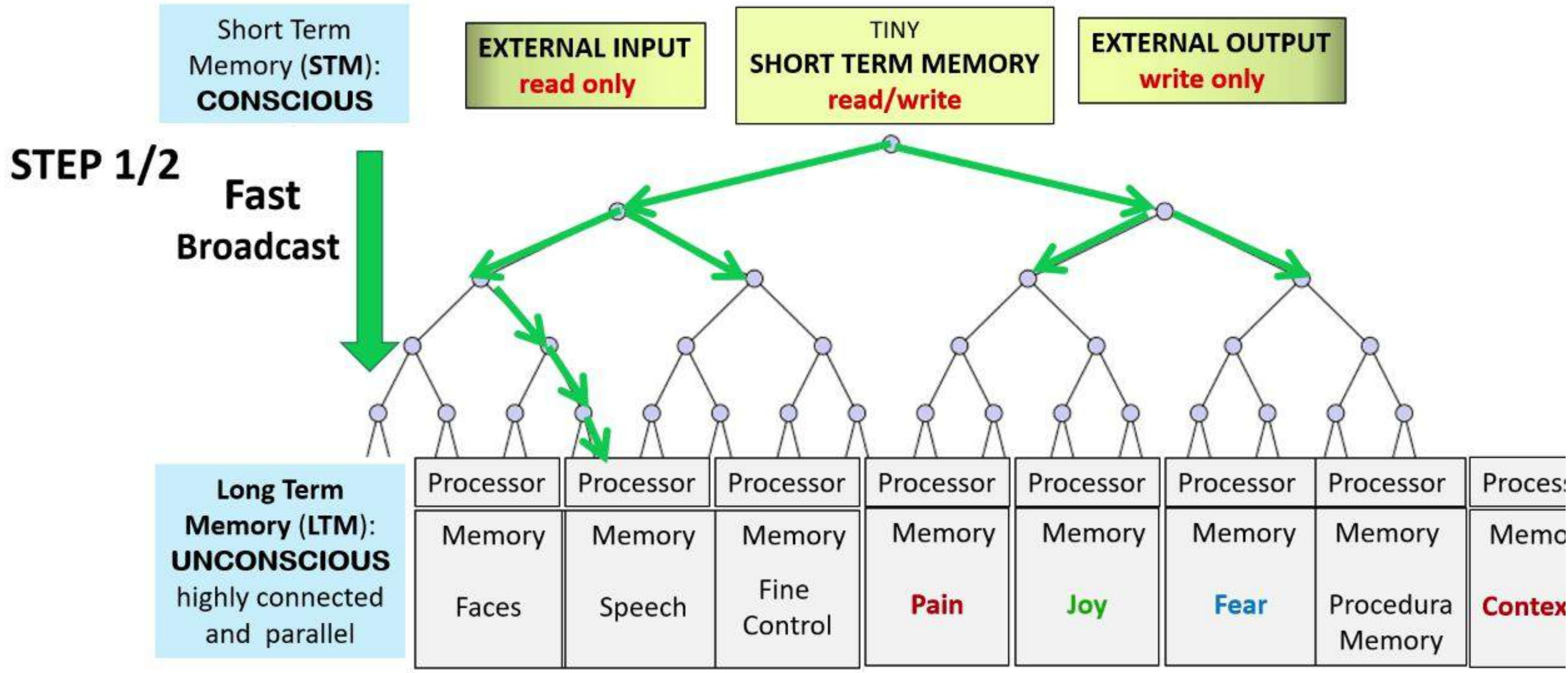
Central Executive function is performed by Unconscious Processors



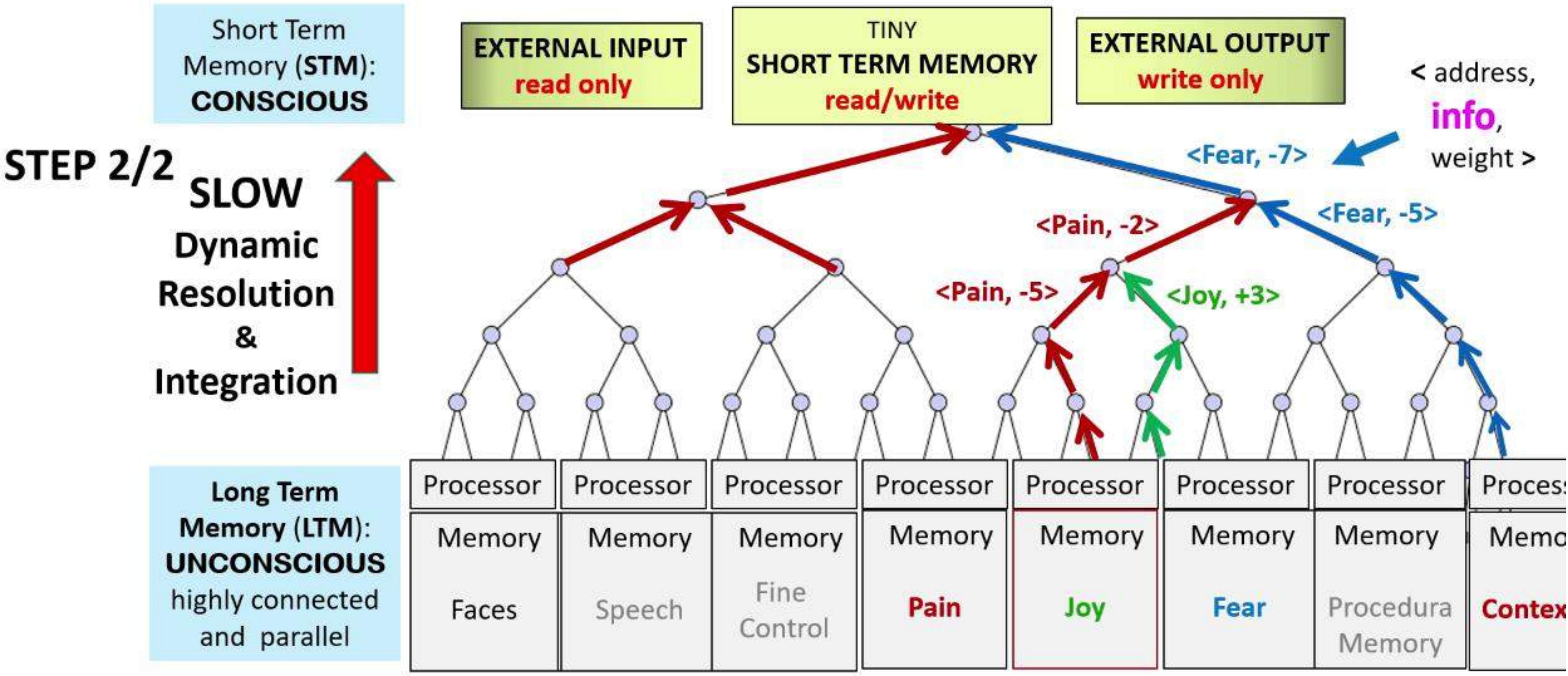
Long Term Memory (**LTM**): **UNCONSCIOUS** highly connected and parallel

Processor	Processor	Processor	Processor	Processor	Processor	Processor	Proces
Memory **Faces**	Memory **Speech**	Memory **Fine Control**	Memory **Fear**	Memory **Desire**	Memory **Embar rassment**	Memory **Declarativ Memory Creation**	Memo **Contex**

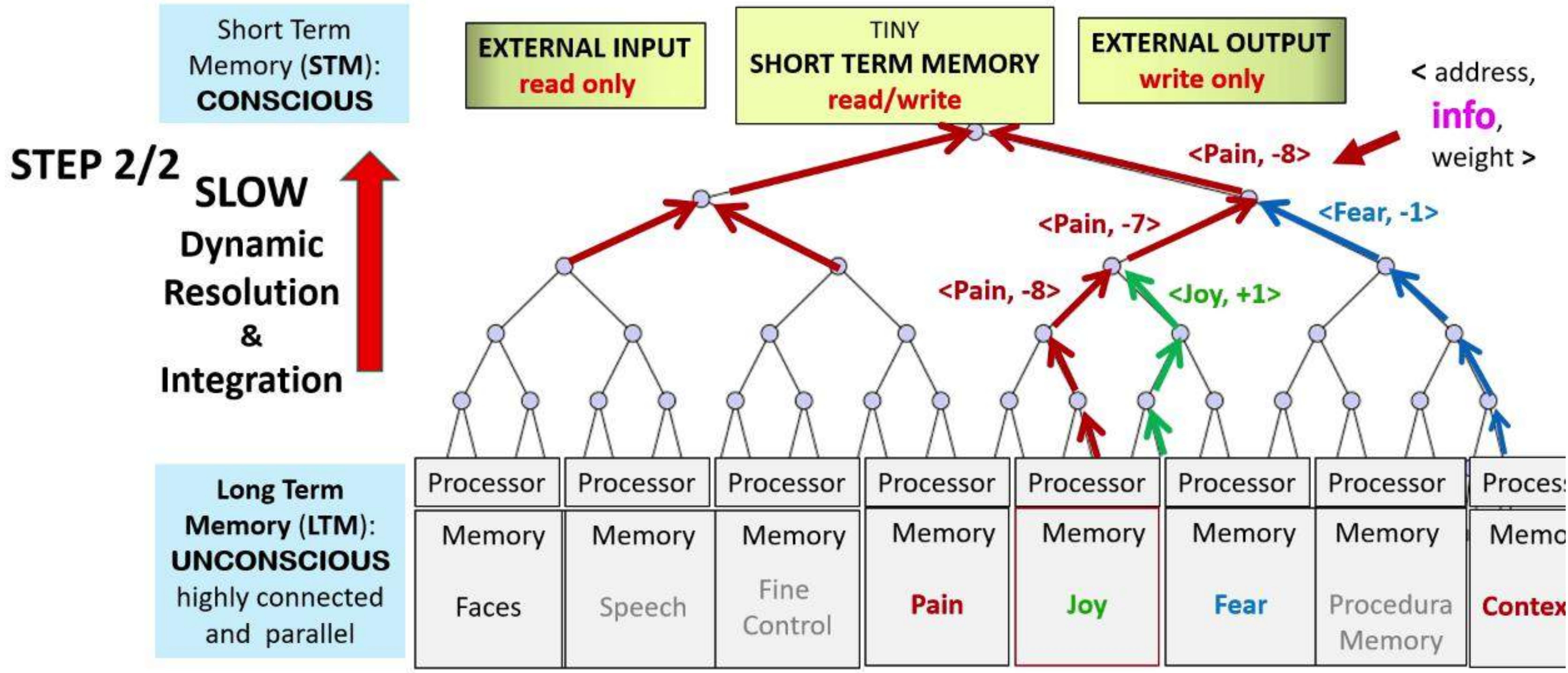
THE CONSCIOUS TM DYNAMICS



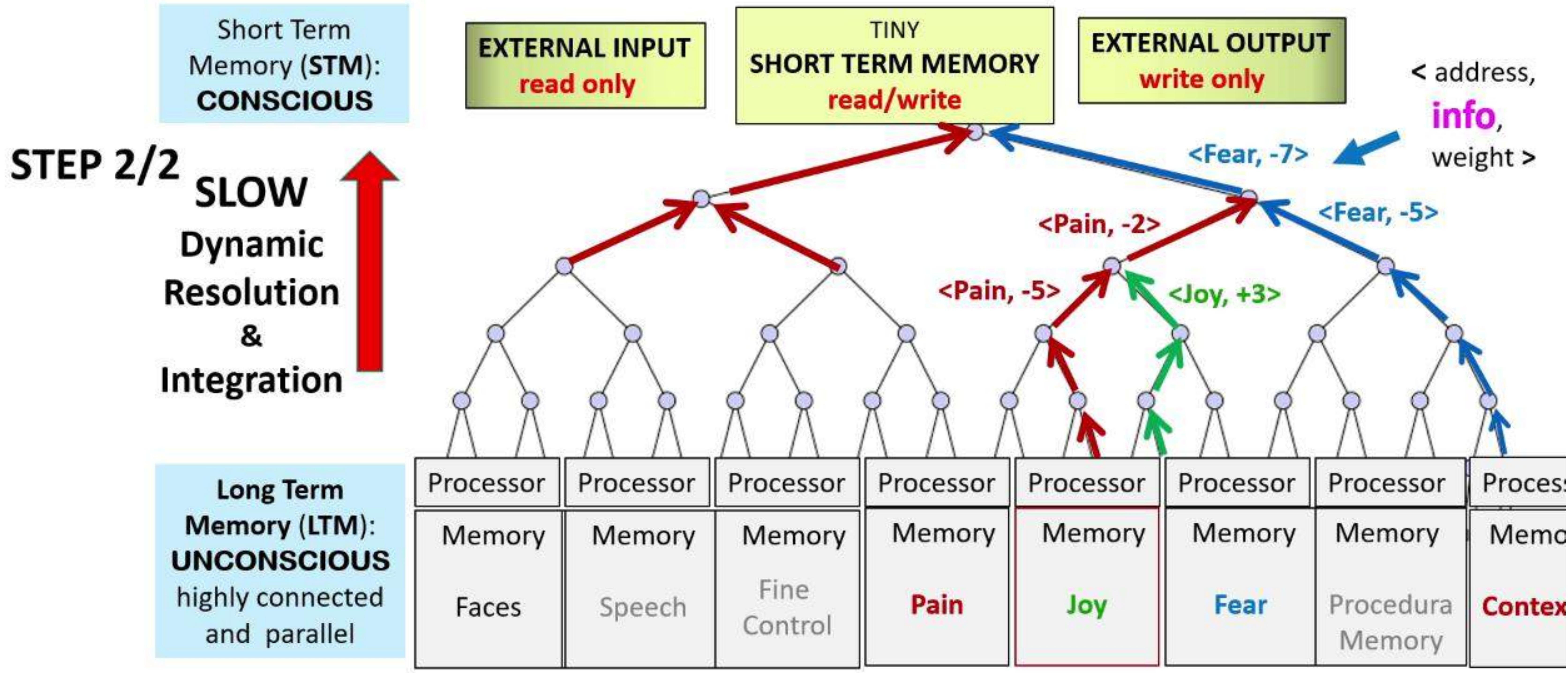
THE CONSCIOUS TM DYNAMICS



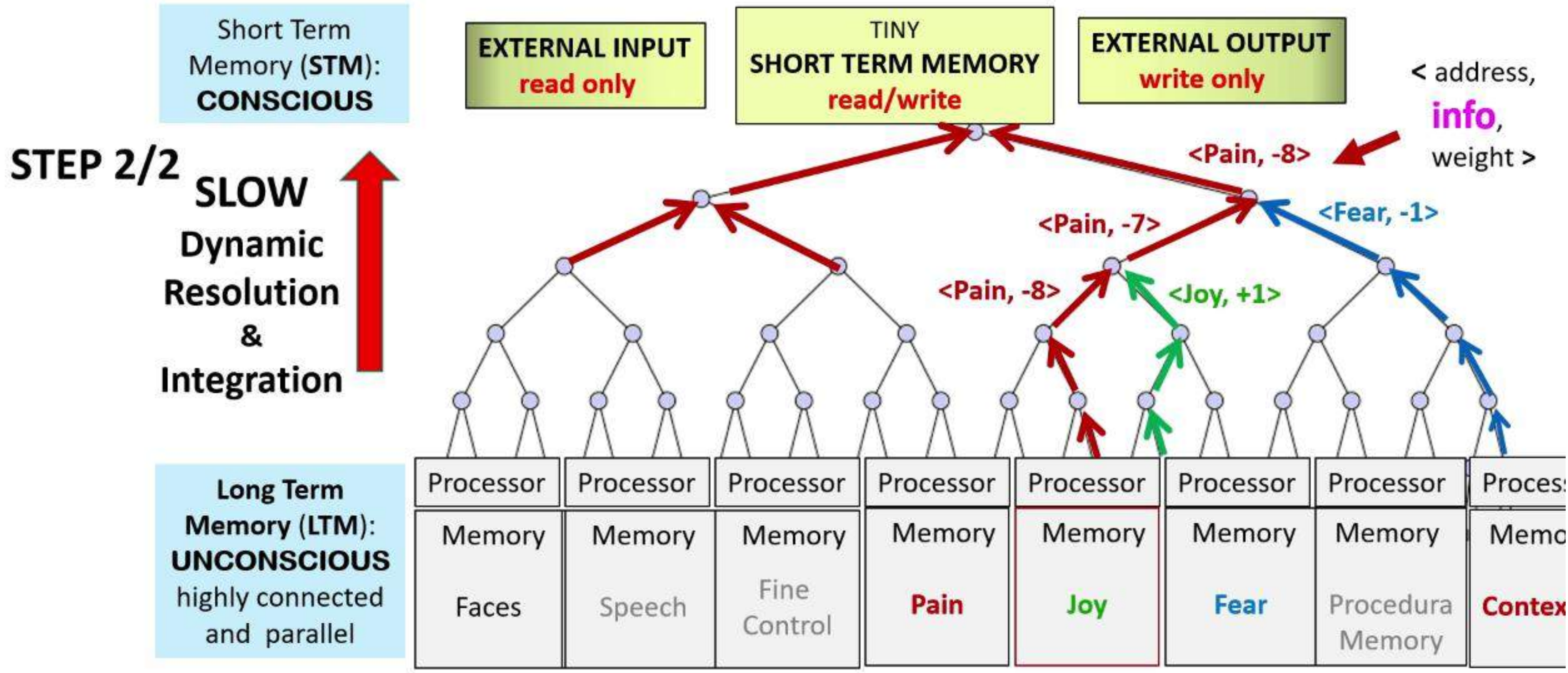
THE CONSCIOUS TM DYNAMICS



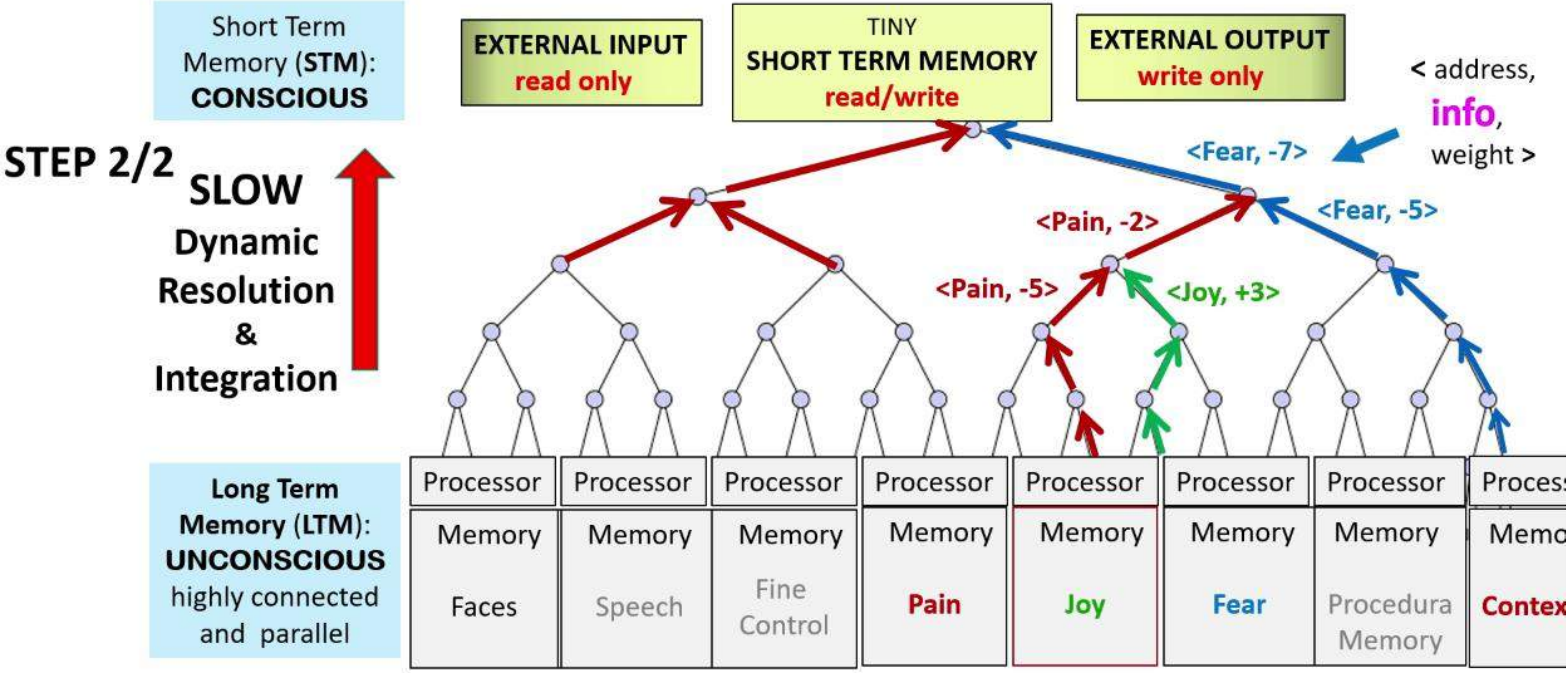
THE CONSCIOUS TM DYNAMICS



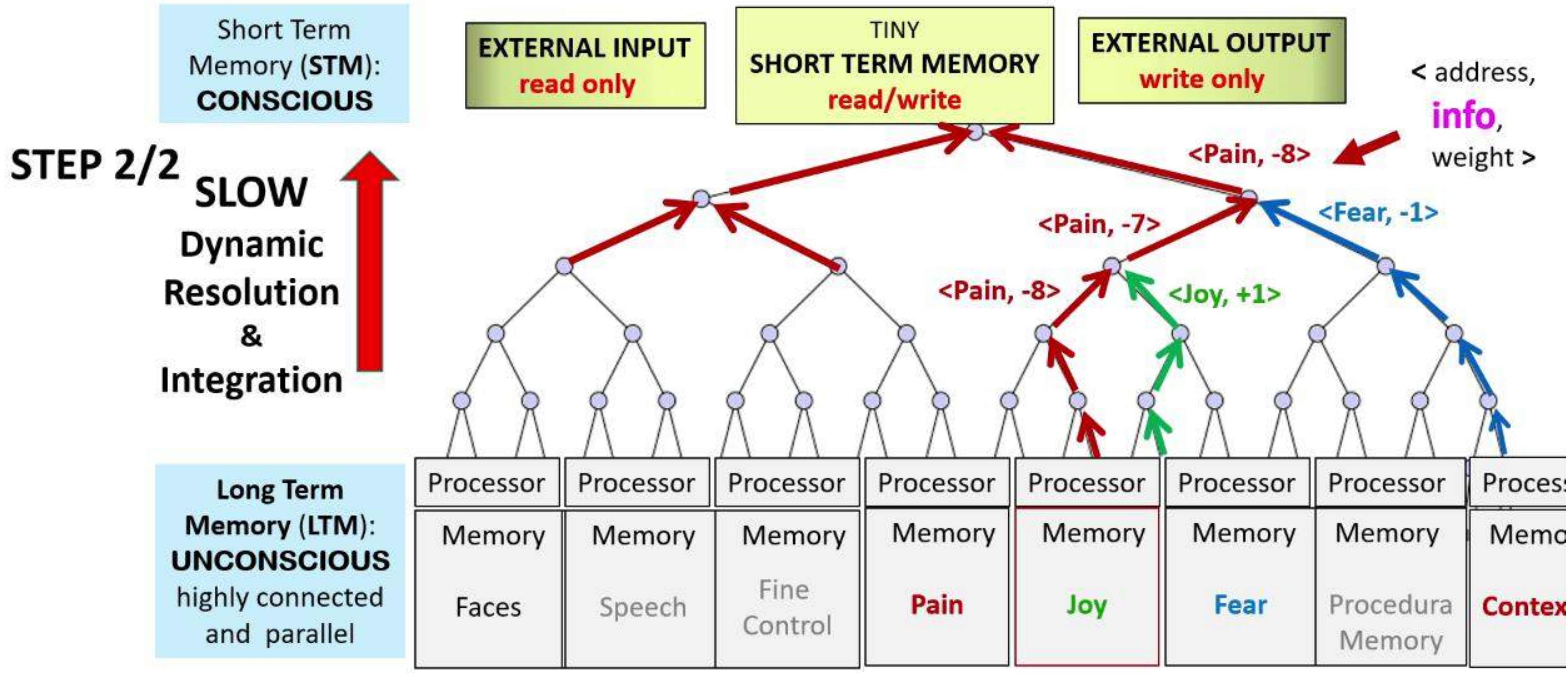
THE CONSCIOUS TM DYNAMICS



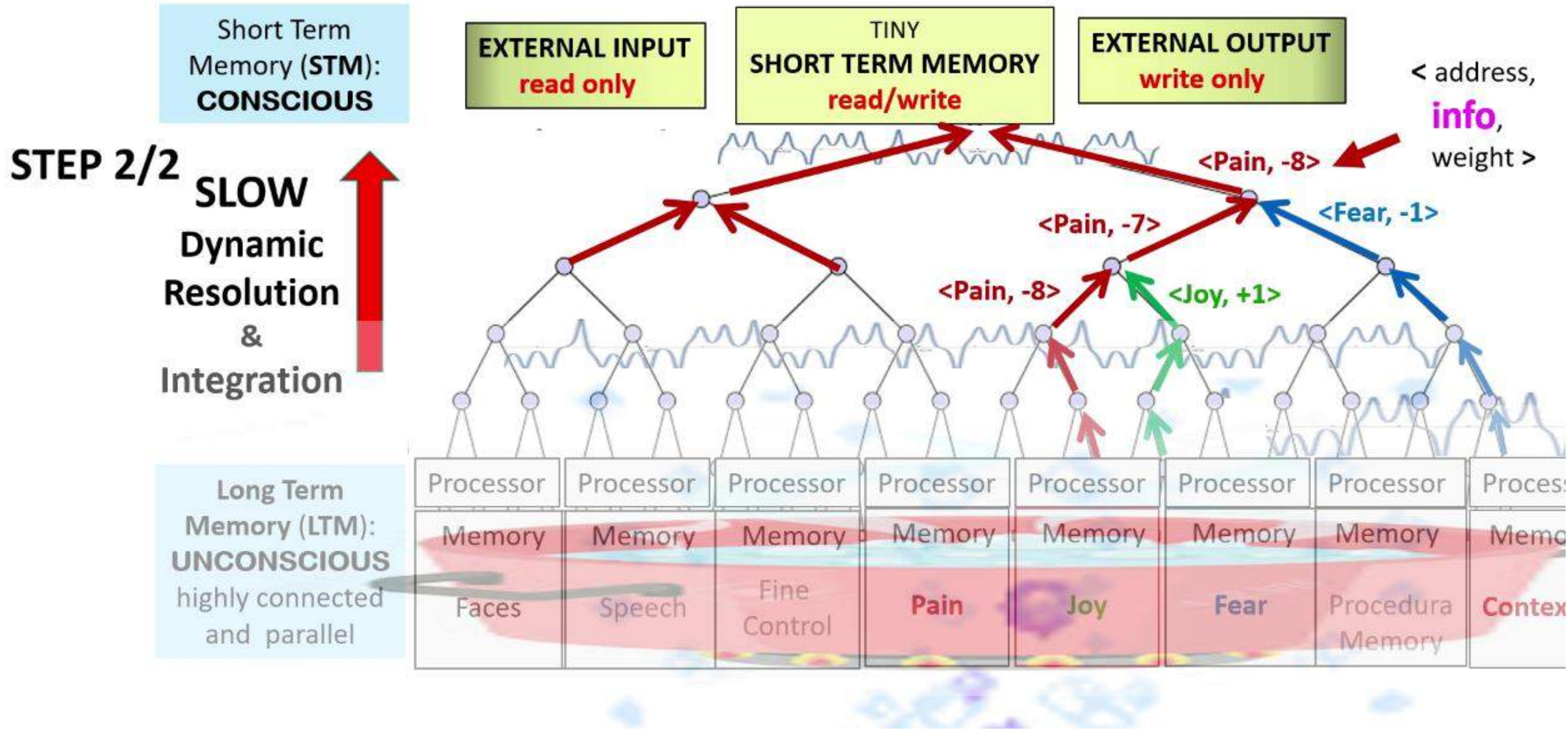
THE CONSCIOUS TM DYNAMICS



THE CONSCIOUS TM DYNAMICS

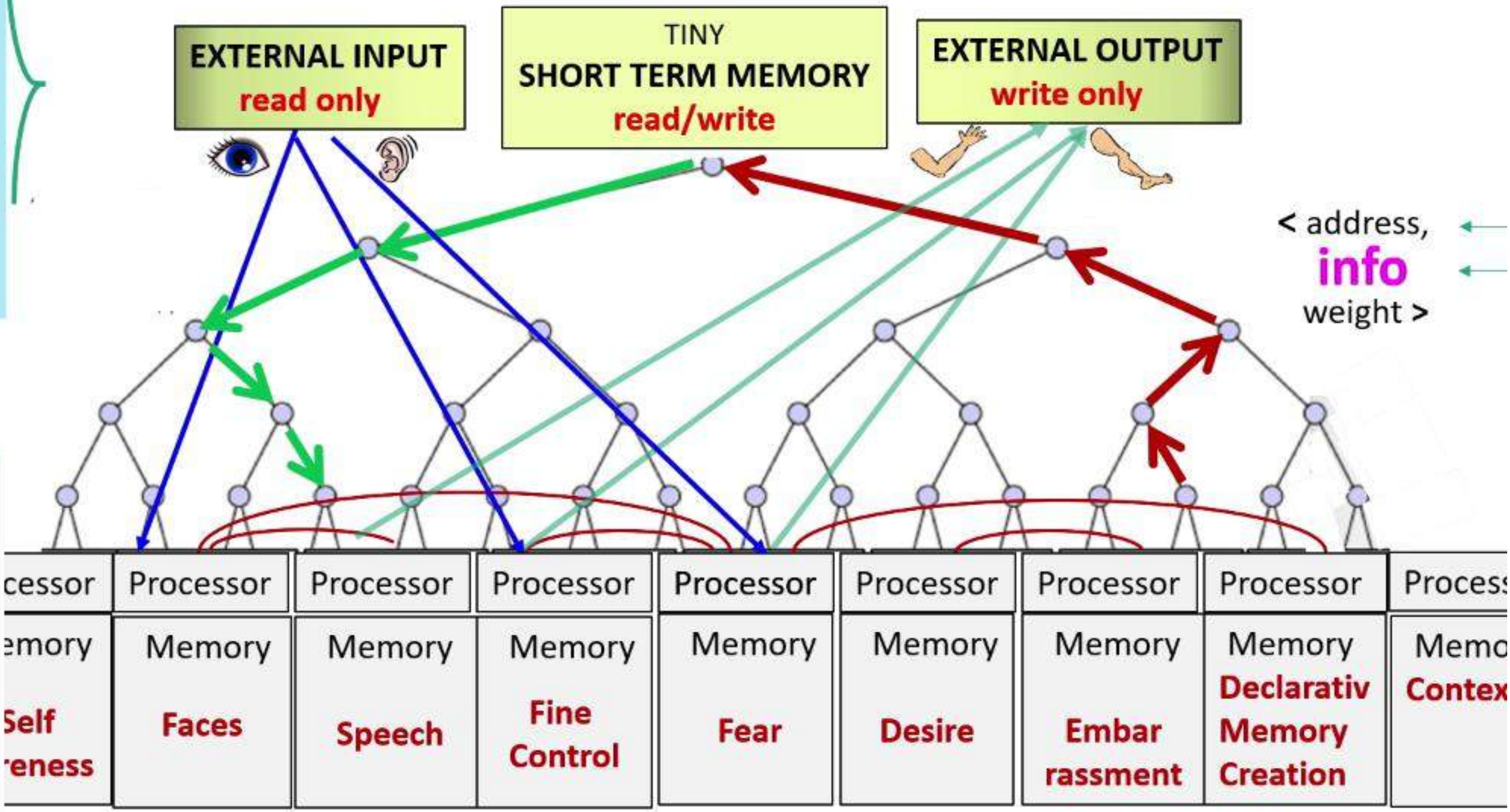


THE CONSCIOUS TM DYNAMICS



THE CONSCIOUS TM IN TOTO

External Input is readable and external output is writable – by all unconscious processors that can use such a link.



Long Term Memory is Enormous. Its processors are unconscious, specialized, parallel and a bit connected

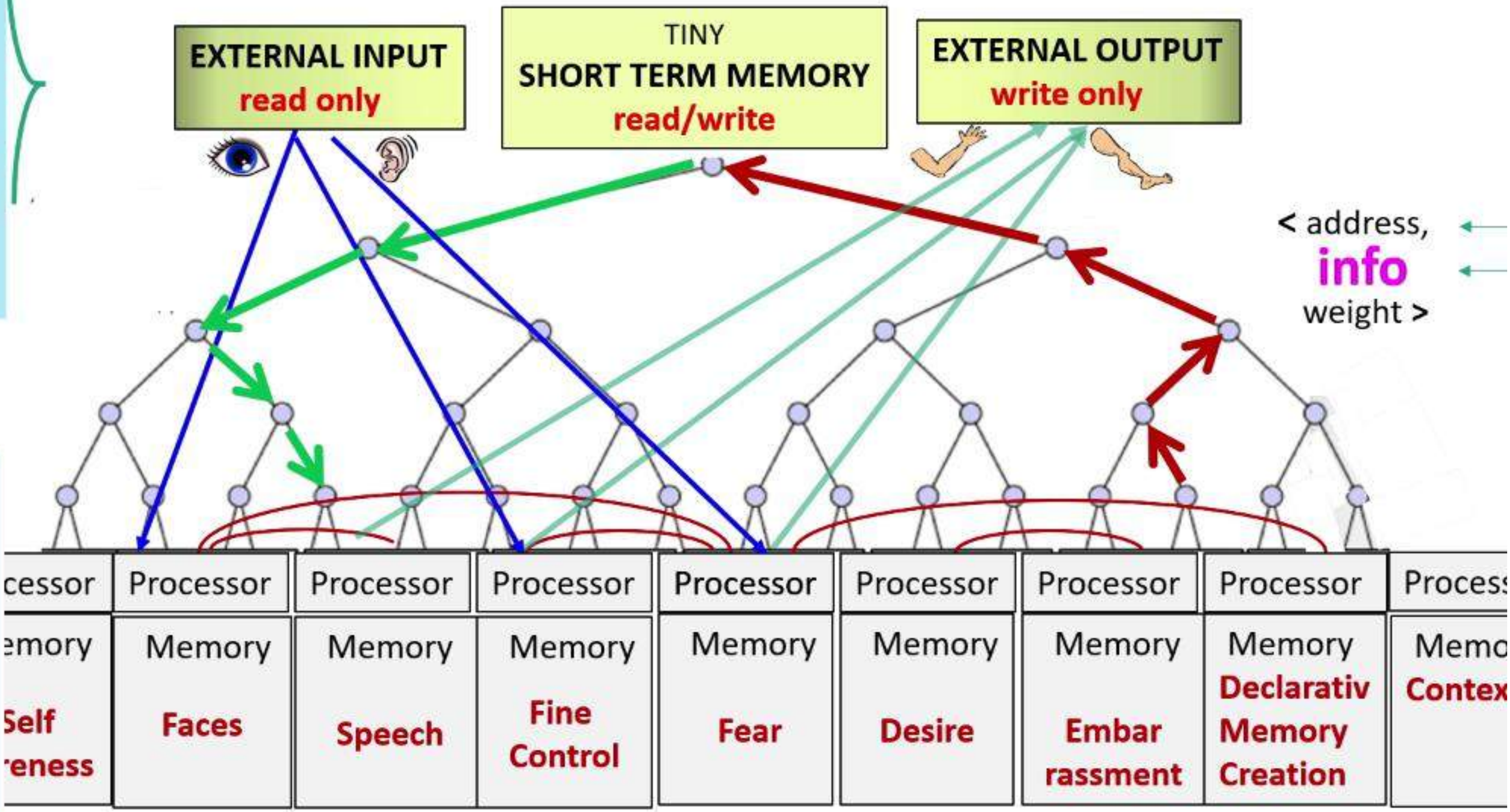
THE CONSCIOUS TURING MACHINE

Details: How processors choose weights

- **A pain process**: In the case of a scraped knee, its **|weight|** is proportional to the number of "nociceptive fibers" that fire and the frequency of their firing. In the case of "fibromyalgia", the **|weight|** is proportional to the number and firing frequency of "brain" cells devoted to pain. The sign of pain **weight** is negative.
- **A context process**: It has a relatively high fixed **weight**, high enough to keep its **info (the scene gist)** on stage except when concentrated attention is required for the task at hand… to contemplate a next move in chess, an experiment to be performed, the proof of a theorem, or the heart-grabbing sound of Louis Armstrong.
- **A task that has been put off**: Its **weight** grows as a function of its importance and the length of time it has been put off.

THE CONSCIOUS TM IN TOTO

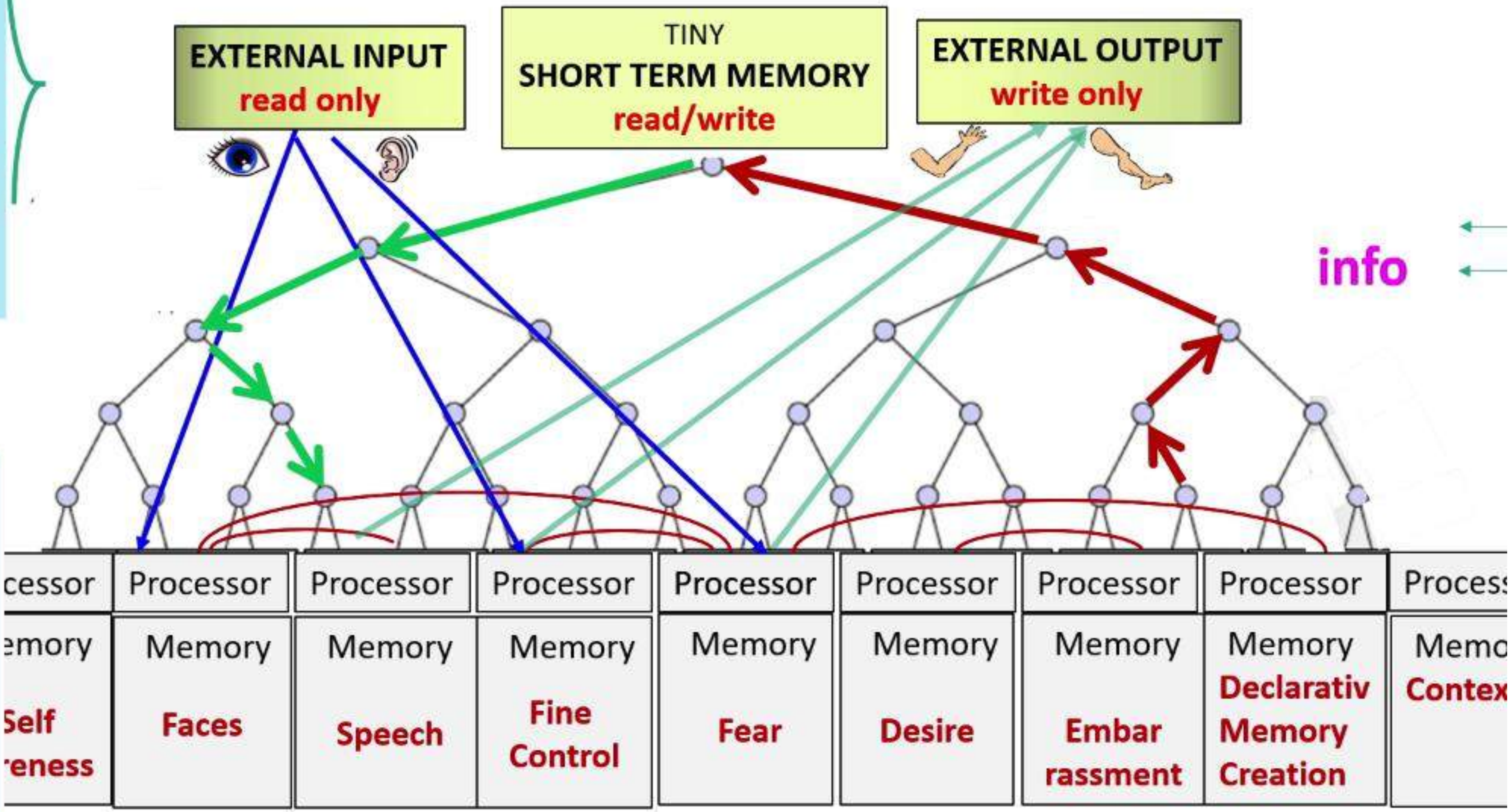
External Input is readable and external output is writable – by all unconscious processors that can use such a link.



Long Term Memory is Enormous. Its processors are unconscious, specialized, parallel and a bit connected

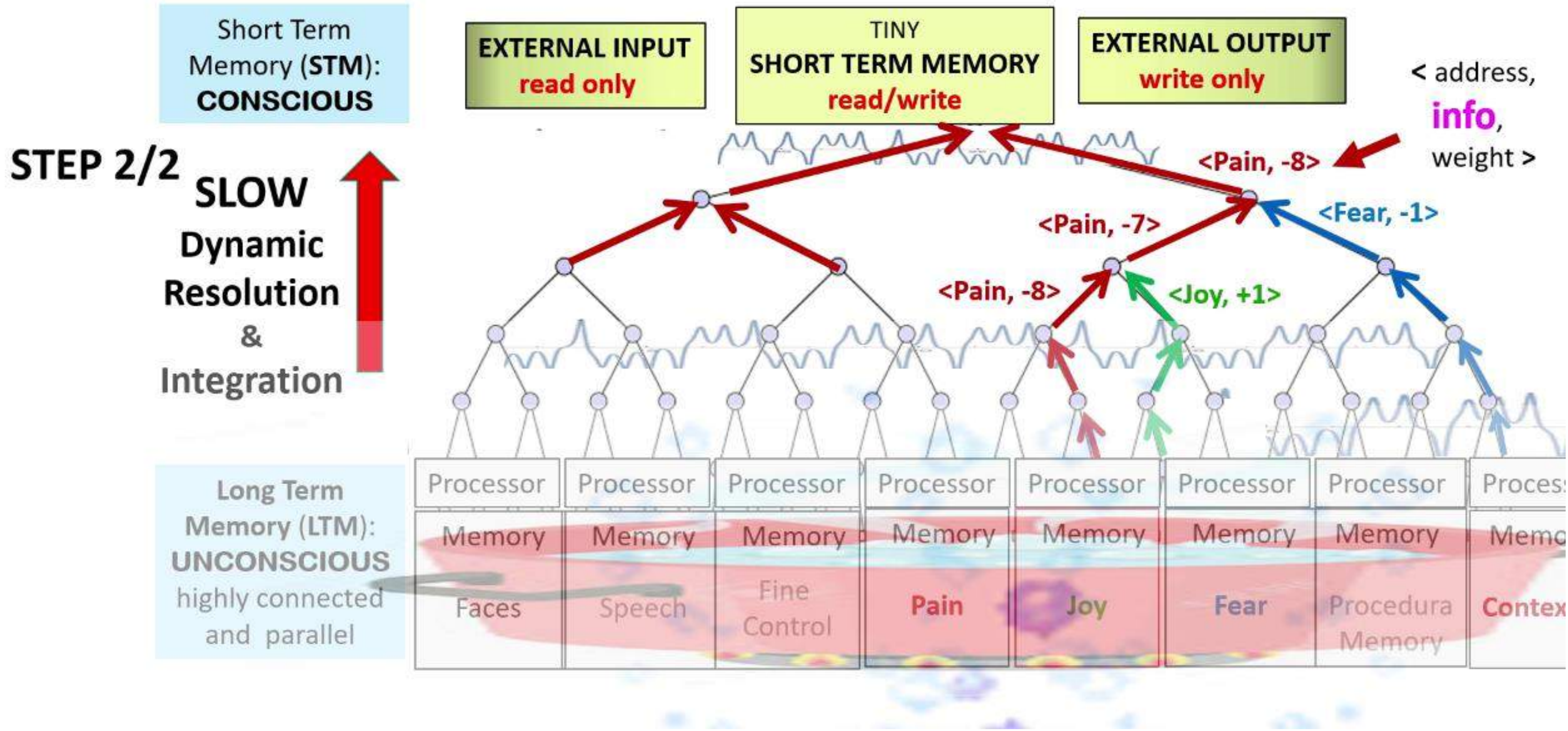
THE CONSCIOUS TM IN TOTO

External Input is readable and external output is writable – by all unconscious processors that can use such a link.



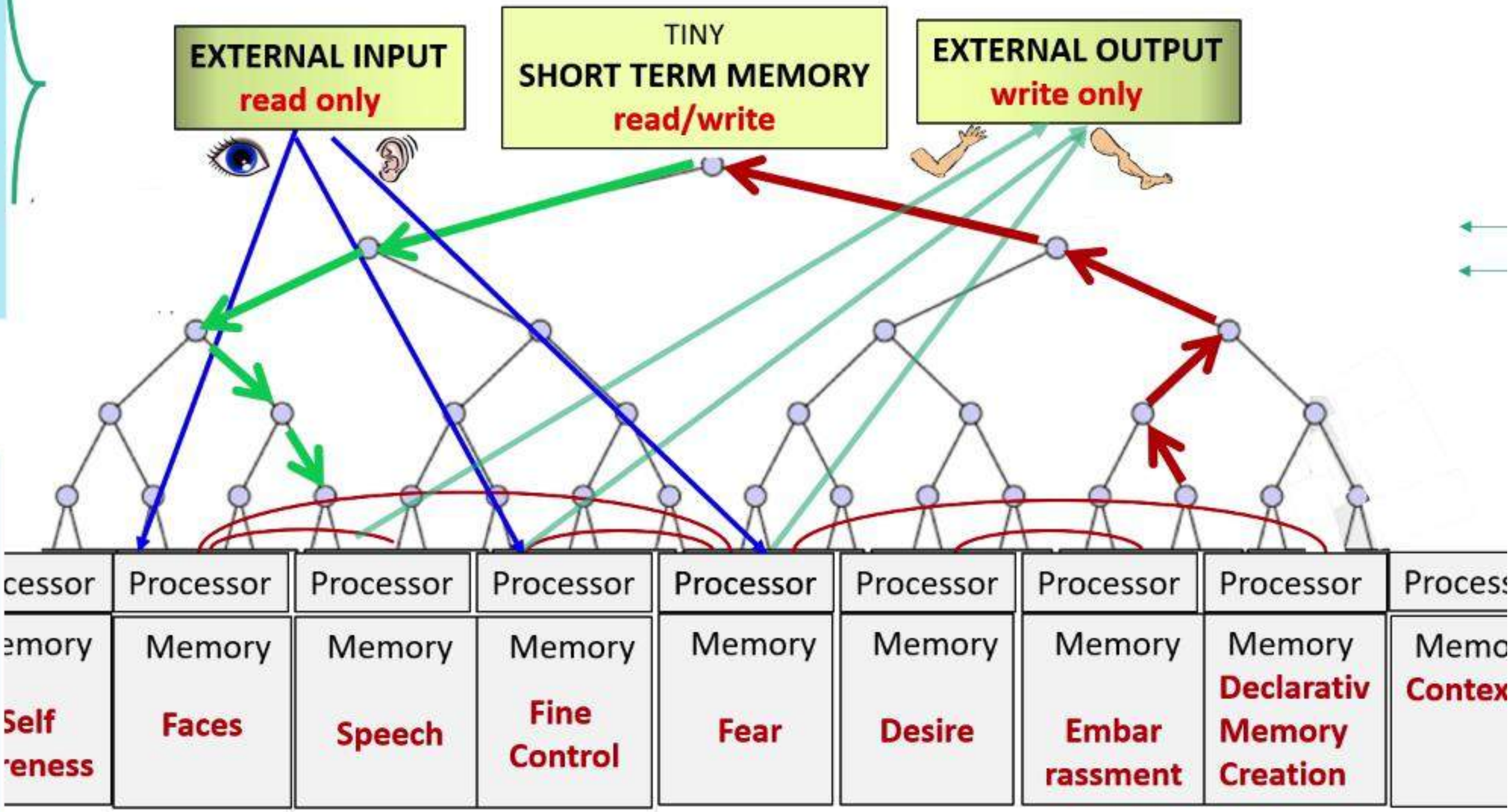
Long Term Memory is Enormous. Its processors are unconscious, specialized, parallel and a bit connected

THE CONSCIOUS TM DYNAMICS



THE CONSCIOUS TM IN TOTO

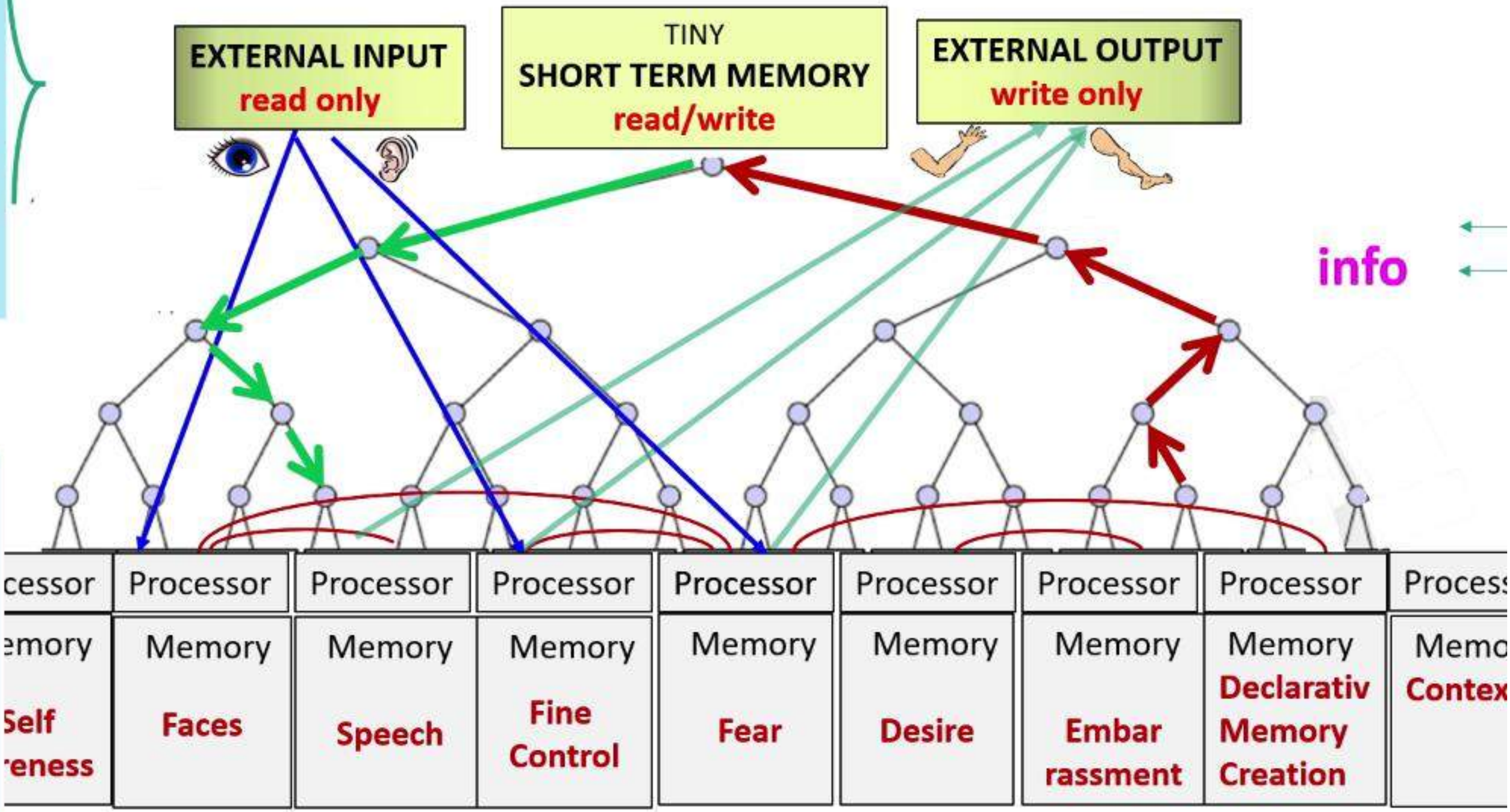
External Input is readable and external output is writable – by all unconscious processors that can use such a link.



Long Term Memory is Enormous. Its processors are unconscious, specialized, parallel and a bit connected

THE CONSCIOUS TM IN TOTO

External Input is readable and external output is writable – by all unconscious processors that can use such a link.



Long Term Memory is Enormous. Its processors are unconscious, specialized, parallel and a bit connected

THE CONSCIOUS TURING MACHINE

Details: How processors choose weights

- **A pain process**: In the case of a scraped knee, its **|weight|** is proportional to the number of "nociceptive fibers" that fire and the frequency of their firing. In the case of "fibromyalgia", the **|weight|** is proportional to the number and firing frequency of "brain" cells devoted to pain. The sign of pain **weight** is negative.
- **A context process**: It has a relatively high fixed **weight**, high enough to keep its **info (the scene gist)** on stage except when concentrated attention is required for the task at hand... to contemplate a next move in chess, an experiment to be performed, the proof of a theorem, or the heart-grabbing sound of Louis Armstrong.
- **A task that has been put off**: Its **weight** grows as a function of its importance and the length of time it has been put off.

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

- There is a threshold for allowing processes into **STM**.
  If set too low, the bubbling of ideas can produce **Mania**.
  If set too high, the absence of ideas produces **Depression**.

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

- There is a threshold for allowing processes into **STM**.
  If set too low, the bubbling of ideas can produce **Mania**.
  If set too high, the absence of ideas produces **Depression**.
- When processes enter **STM**, they slowly lose weight. They slowly regain their weight after they exit **STM**. This can lead to cycling among processes of roughly equal weight.[1]

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

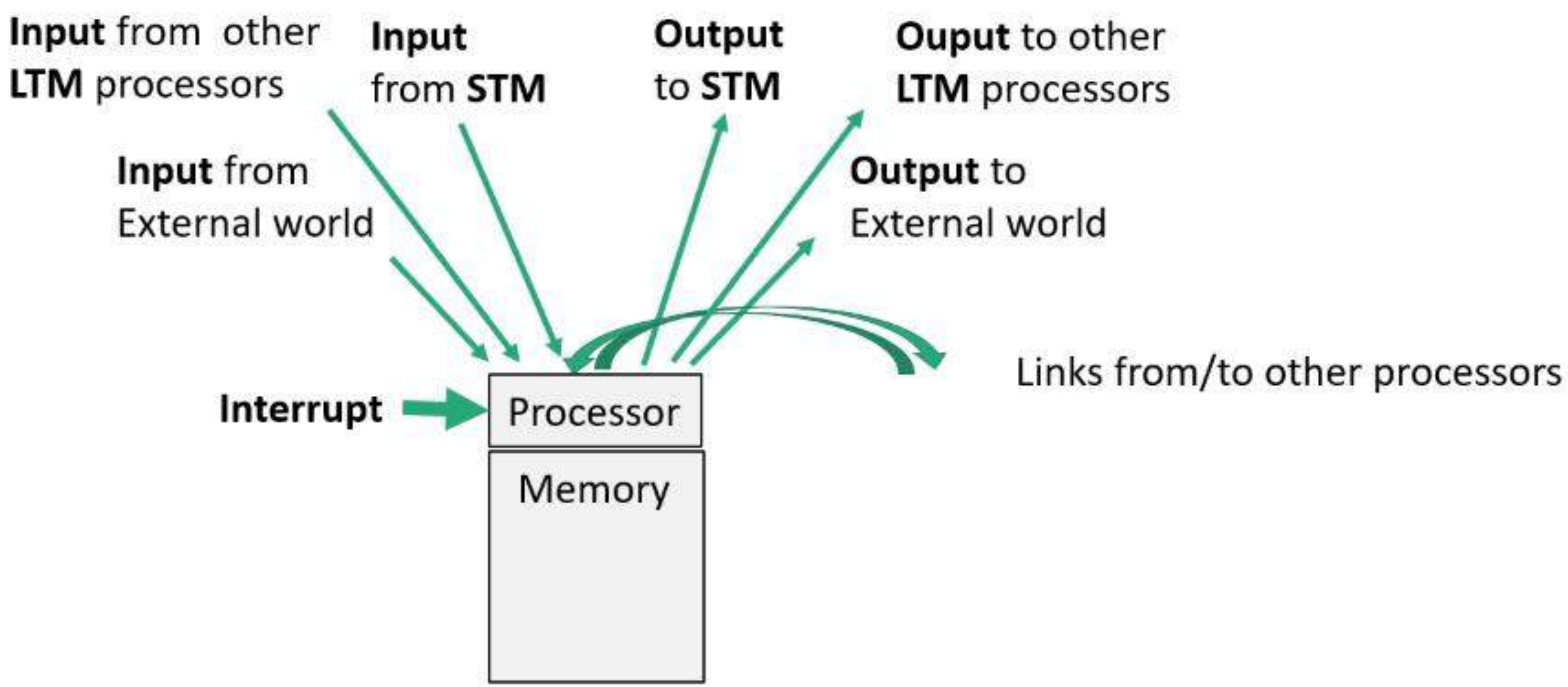
- There is a threshold for allowing processes into **STM**.
  If set too low, the bubbling of ideas can produce **Mania**.
  If set too high, the absence of ideas produces **Depression**.
- When processes enter **STM**, they slowly lose weight. They slowly regain their weight after they exit **STM**. This can lead to cycling among processes of roughly equal weight.[1]
- **LTM** processor **A** links up to **B** when **A** answers **B's** call.
  Linking enables conscious processing to become unconscious.

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

- There is a threshold for allowing processes into **STM**.
  If set too low, the bubbling of ideas can produce **Mania**.
  If set too high, the absence of ideas produces **Depression**.
- When processes enter **STM**, they slowly lose weight. They slowly regain their weight after they exit **STM**. This can lead to cycling among processes of roughly equal weight.[1]
- **LTM** processor **A** links up to **B** when **A** answers **B's** call.
  Linking enables conscious processing to become unconscious.

The LTM Processor



THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

Some Unconscious Processors (aka *Sleeping Experts*) give too much weight to their information, some too little.

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

Some Unconscious Processors (aka *Sleeping Experts*)
give too much weight to their information, some too little.

Avrim Blum's "Sleeping Experts" Theorem presents an optimal algorithm to *dynamically refine* weights assigned to information by *competing processors.*

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

Some Unconscious Processors (aka *Sleeping Experts*) give too much weight to their information, some too little.

Avrim Blum's "Sleeping Experts" Theorem presents an optimal algorithm to *dynamically refine* weights assigned to information by *competing processors.*

- **The algorithm incorporates feedback from experience.**

THE CONSCIOUS TURING MACHINE

Details of the Dynamics:

Some Unconscious Processors (aka *Sleeping Experts*) give too much weight to their information, some too little.

Avrim Blum's "Sleeping Experts" Theorem presents an optimal algorithm to *dynamically refine* weights assigned to information by *competing processors.*

- **The algorithm incorporates feedback from experience.**

See Theorem 5.21 in the book: ***Foundations of Data Science*** by **Avrim Blum, John Hopcroft, and Ravikandran Kannan**

CONSCIOUSNESS

Recall that *Consciousness* in the model is the content of Short Term Memory (STM).

CONSCIOUSNESS

Recall that *Consciousness* in the model is the content of Short Term Memory (STM).

How reasonable is this definition of consciousness?

1. **all LTM processors are focused on what's on stage,**

CONSCIOUSNESS

Recall that *Consciousness* in the model is the content of Short Term Memory (STM).

How reasonable is this definition of consciousness?

1. **all LTM processors are focused on what's on stage,**
2. **The stage activity can be persistent.**

CONSCIOUSNESS

Recall that *Consciousness* in the model is the content of Short Term Memory (STM).

How reasonable is this definition of consciousness?

1. **all LTM processors are focused on what's on stage,**
2. **The stage activity can be persistent.**

Many questions:

Why is **short term memory** so tiny?   What is a **chunk**?

CONSCIOUSNESS

Recall that *Consciousness* in the model is the content of Short Term Memory (STM).

How reasonable is this definition of consciousness?

1. **all LTM processors are focused on what's on stage,**
2. **The stage activity can be persistent.**

Many questions:

Why is **short term memory** so tiny? What is a **chunk**?

< address,
info,
weight >

Why am I interested in consciousness?

- It is obviously **useful to humans**.

Why am I interested in consciousness?

- It is obviously **useful to humans**.
- It's focuses **LTM** processors on creating the current **best interpretation of the world**.

Why am I interested in consciousness?

- It is obviously **useful to humans**.
- It's focuses **LTM** processors on creating the current **best interpretation of the world**.
- It's a **checker** on that interpretation.

Why am I interested in consciousness?

- It is obviously **useful to humans**.
- It's focuses **LTM** processors on creating the current **best interpretation of the world**.
- It's a **checker** on that interpretation.
- It gives the entity the **ability to solve unanticipated problems**, to deal with a complex world using all the tools at its disposal.

The conscious is not privy to the workings of the unconscious:

Example 1: **What's her name?**

You recall something about her, perhaps where you first met ↗

It is broadcast ↘. Something about what she does ↗

It is broadcast ↘. You recall that her name begins with T ↗

It is broadcast ↘. All these recollections come ↗ from **LTM** to **STM.**

A half hour later her name comes to you ↗, surfacing from the unconscious **LTM**, which has been thinking, searching.

LTM is unconscious: you don't really know how or where **LTM** found her name.

THE CONSCIOUS TURING MACHINE

Example 2: how this might work

Oliver Sacks and the bull

- Oliver Sacks, while hiking on a mountain, chances upon a bull. On seeing it, Oliver's entire consciousness **(STM)** shrieks **fear.**
- All **LTM** processors (audience) concentrate on **fear.**
- **Fight** and **flight** processors vie for the stage. **Flight** wins out…
- From stage, **Flight** signals legs: **Turn. Run!**
- Had all processors on stage remained fixed on **fear** and not given access to the stage, Oliver might have **frozen**.

THE CONSCIOUS TURING MACHINE

Example 3: Recollections and Dreams:

Imagine the entrance to your home. It may help to articulate details like stairs, porch, plants. **Compare that memory to the real thing. Memory is incomparably hazier than the real thing.** **Compare your dreams to the real thing. If you have ever had a lucid dream, then you know how real dreams can be.** **Your brain is capable of generating exquisitely detailed images.**

Why does your brain generate hazy memories when it can generate much more realistic ones?

Possible answer: so you don't confuse memories with the real thing.

Why do you forget your dreams (unless you work to keep them)?

Possible answer: so you don't confuse dreams with the real thing.

THE CONSCIOUS TURING MACHINE

Example 4: Grasping the (well-understood) proof of a Theorem

A proof is in general a directed graph. In the example below, we show just its acyclic spanning tree (leaving out links that obscure the proof):

1. The root of the tree is a statement of the Theorem.
2. The next level of the tree is a very high-level idea of the proof, succinctly and informally stated.
3. This proof idea is expanded at the next level into several sub-ideas. These sub-ideas are children of the level 2 parent idea. They are siblings, each of which includes info about its relation to its other siblings and its parent.

:

Each node of the tree is therefore an idea, with its own children.

:

At the bottom, the leaves of this tree are basic building blocks, typically Lemmas, possibly related elementary statements like A, B, and A => B.

THE CONSCIOUS TURING MACHINE

Example 4 continued:

1: **Theorem**: sqrt(2) is not rational.

2: **Proof**: Assume to contrary that sqrt(2) is rational number a/b.

3: (**child 1**: a, b are rel prime positive integers & **child 2**: sqrt(2) = a/b).

4: **child 1.1:** for all prime p, p cannot divide both a and b.

4: **child 2.1:** squaring both sides, $2 = a^2/b^2$.

5: **child 2.1.1**: multiplying both sides by b^2, $2b^2 = a^2$.

6: **child 2.1.1.1**: 2 divides LHSE ($2b^2$), hence 2 divides RHSE (a^2).

7. (**child 2.1.1.1.1**: 2 is prime & **child 2.1.1.1.2**: 2 divides a) ⇒ (**child 2.1.1.1.3**: 4 divides a^2).

The partial tree created above is far from unique. Anyone who understands the proof can have his/her own structure and sentences, and can use that to traverse from any node to any of its descendants.

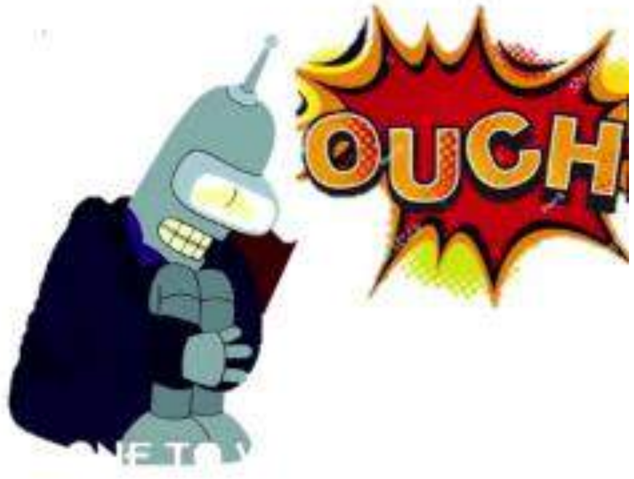
The Hard Problem

Consider the Hard Problem for the special case of pain.

The Hard Problem

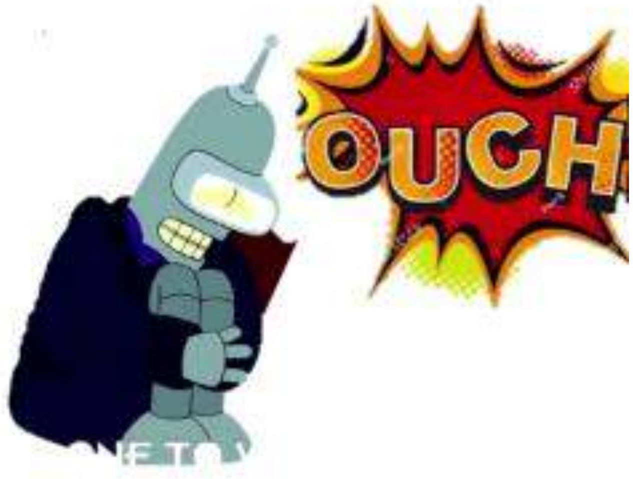
Consider the Hard Problem for the special case of pain. **Why?**

The Hard Problem



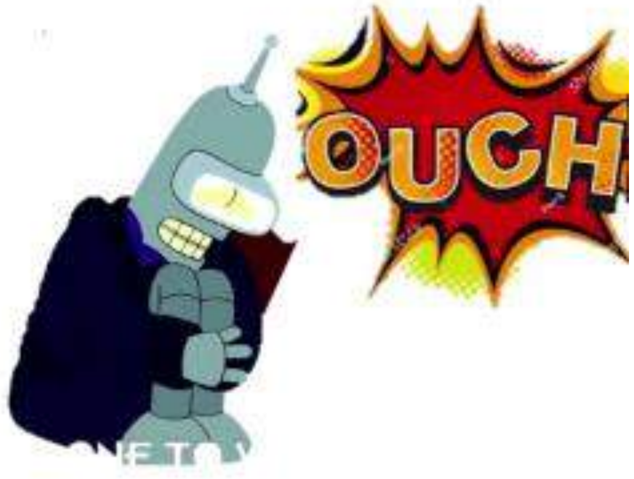
Consider the Hard Problem for the special case of pain.

The Hard Problem



Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

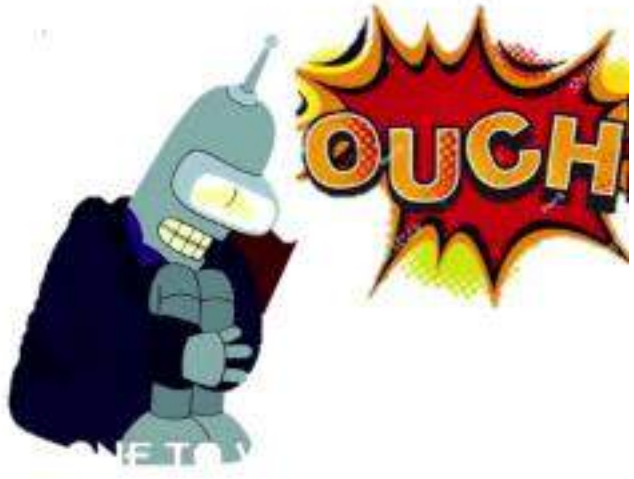
The Hard Problem



Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations.

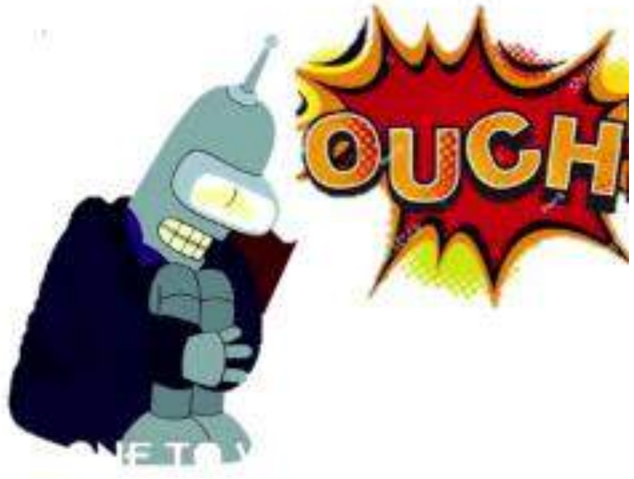
The Hard Problem



Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations. Here are suggestions that **don't work** -- as attested to by asymbolics type 2.

The Hard Problem



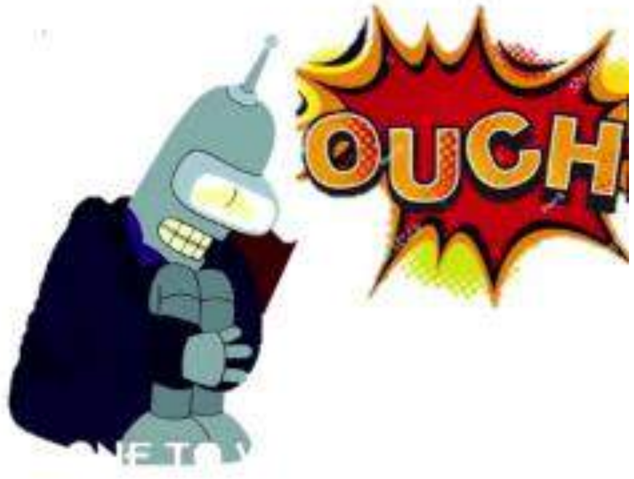
Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations. Here are suggestions that **don't work**
-- as attested to by asymbolics type 2.

Pain might arise from observing unconscious reactions such as

- **grimacing, crying out, and such.**

The Hard Problem



Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations. Here are suggestions that **don't work** -- as attested to by asymbolics type 2.

Pain might arise from observing unconscious reactions such as

- **grimacing, crying out, and such.**
- response to painful situations such as a finger pulling away from a flame (Cat on a Hot Tin Roof).

The Hard Problem

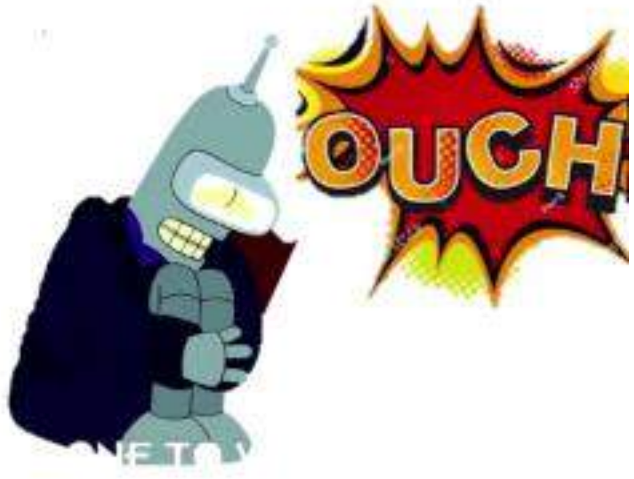
Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations. Here are suggestions that **don't work** -- as attested to by asymbolics type 2.

Pain might arise from observing unconscious reactions such as

- **grimacing, crying out, and such.**
- response to painful situations such as a finger pulling away from a flame (Cat on a Hot Tin Roof).
- sweat and increased heart rate (or its equivalent).

The Hard Problem



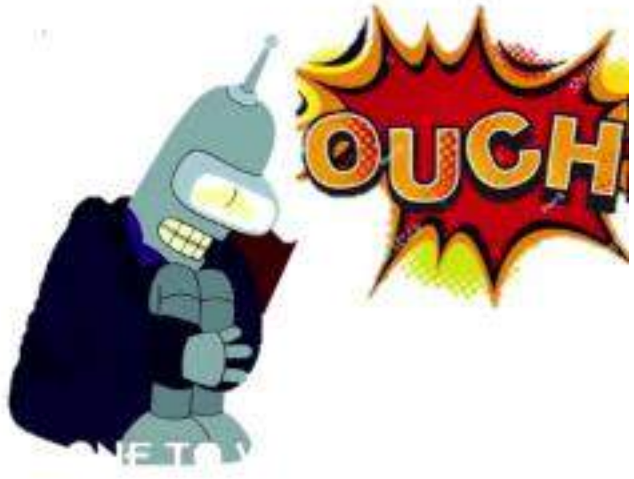
Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations. Here are suggestions that **don't work** -- as attested to by asymbolics type 2.

Pain might arise from observing unconscious reactions such as

- **grimacing, crying out, and such.**
- response to painful situations such as a finger pulling away from a flame (Cat on a Hot Tin Roof).
- sweat and increased heart rate (or its equivalent).
- muscles that vibrate involuntarily. Limits on skeletal movements.

The Hard Problem



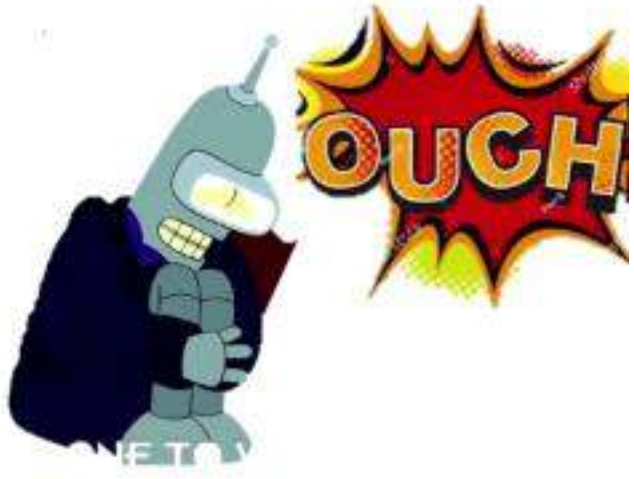
Consider the Hard Problem for the special case of pain.
How might the Conscious Turing Machine experience pain?

We tried many explanations. Here are suggestions that **don't work** -- as attested to by asymbolics type 2.

Pain might arise from observing unconscious reactions such as

- **grimacing, crying out, and such.**
- response to painful situations such as a finger pulling away from a flame (Cat on a Hot Tin Roof).
- sweat and increased heart rate (or its equivalent).
- muscles that vibrate involuntarily. Limits on skeletal movements.
- Nausea. Vomiting. Peeing.

The Hard Problem



Here are our suggestions for **extreme** pain:

The Hard Problem

Here are our suggestions for **extreme** pain:

1. **Broadcasts. Extreme pain** Is an actor that takes over all Short Term Memory. It prevents all other actors (processors) from reaching the stage. Pain messages - and only pain messages - are broadcast. Every processor knows of the pain.

The Hard Problem

Here are our suggestions for **extreme** pain:

1. **Broadcasts. Extreme pain** Is an actor that takes over all Short Term Memory. It prevents all other actors (processors) from reaching the stage. Pain messages - and only pain messages - are broadcast. Every processor knows of the pain.

   Confirmation

The Hard Problem

Here are our suggestions for **extreme** pain:

1. **Broadcasts. Extreme pain** Is an actor that takes over all Short Term Memory. It prevents all other actors (processors) from reaching the stage. Pain messages - and only pain messages - are broadcast. Every processor knows of the pain.

Confirmation

- Under conditions that produce great pain, asymbolics can think while normals cannot.
- In a Darwinian design, you might expect pain to lead to constructive thinking, but agony actually inhibits constructive thinking. If forces you to rely on your unconscious self.

The Hard Problem

Here are our suggestions for **extreme** pain:

1. **Broadcasts. Extreme pain** Is an actor that takes over all Short Term Memory. It prevents all other actors (processors) from reaching the stage. Pain messages - and only pain messages - are broadcast. Every processor knows of the pain.

   Confirmation

   - Under conditions that produce great pain, asymbolics can think while normals cannot.
   - In a Darwinian design, you might expect pain to lead to constructive thinking, but agony actually inhibits constructive thinking. If forces you to rely on your unconscious self.

But... this does not account for the sudden excruciating pain at the moment you tear a ligament. What does?

The Hard Problem

Here are our suggestions for **sudden extreme** pain:

2. **Interrupts** (as opposed to **broadcasts**). Sudden extreme pain - a finger touching a burning stove – **interrupts all** unconscious processors.

The Instant Shock of PAIN

- When a ligament is torn, the shock of pain is instantaneous. How does this excruciating pain come about?
- We suggest that sudden severe pain interrupts what takes place on stage. The interrupt travels down the tree to **all** unconscious processors, forcing each and every processor to decide on the spot whether to continue what it has been doing (if it has nothing to contribute) or attend to the interrupt.
- For example, when the tear in the ligament interrupts a face recognition processor, it forces the processor to put current work on a stack (whose face was it?), then deal with the interrupt as basic mission: are there faces around? If so, are they friend or foe? Who are they?
- For example, when the tear in the ligament travels down the tree and interrupts a fear processor, it causes the processor to compare its current |weight| (fear of the bull) to the interrupting |weight| (torn ligament), and either continue to deal with the fear (running) or hand control to the interruption (deal with the pain). Weird: when pain forces an interrupt, why doesn't fear opt out?
- As confirmation of the interrupt and automatic response to it, people are known to remember exactly where they were when they tore a ligament. Autobiographical memory got interrupted. It stored the info. The decision to store was not made consciously. That's what autobiographical memory does.

We propose: The sudden interrupt of all processing systems registers as shock.

The Hard Problem

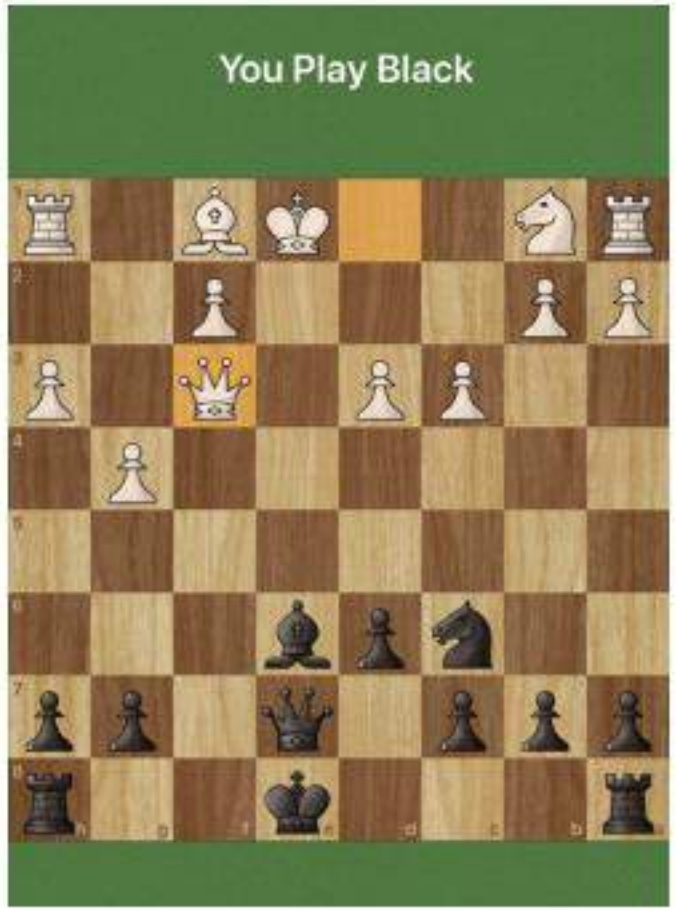
Consider the Hard Problem for the special case of joy
How might the Conscious Turing Machine experience joy?

Here are our suggestions for **extreme** joy:

Free Will

- **Free will** is the ability to compute the consequences of different courses of action and choose accordingly.

The example of chess:

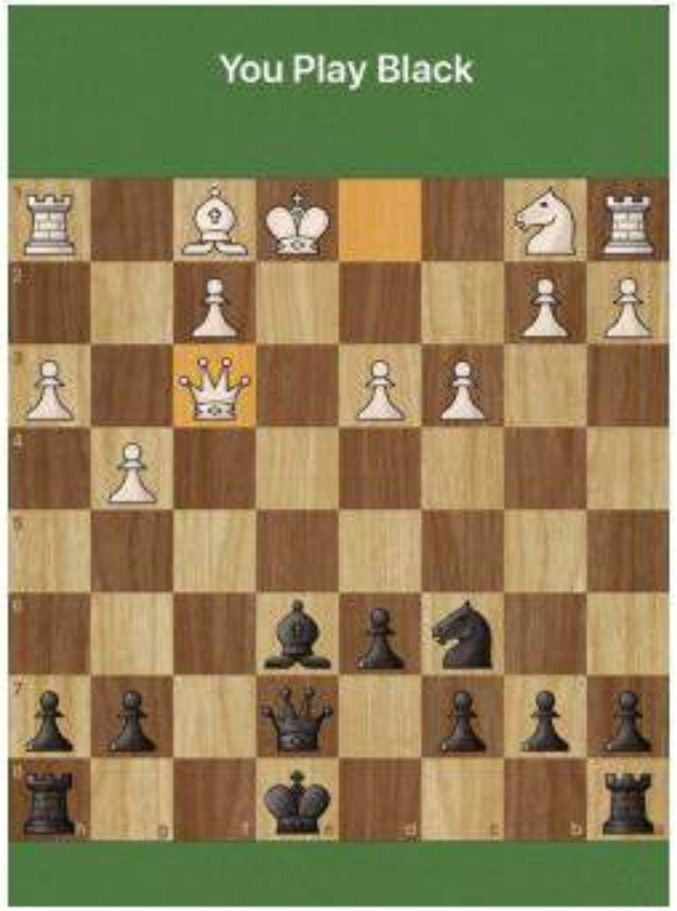


Computation is not instantaneous

Free Will

- **Free will** is the ability to compute the consequences of different courses of action and choose accordingly.

The example of chess:



Computation is not instantaneous

The solution beautifully explained by Stanislas Dehaene in *Consciousness and the Brain,* 2014:

"**Our brain states are clearly not uncaused and do not escape the laws of physics** – nothing does. But our **decisions are genuinely free whenever they are based** on a conscious deliberation that proceeds autonomously, without any impediment, **carefully weighing the pros and cons before committing to a course of action.** When this occurs, we are correct in speaking of a voluntary decision – even if it is, of course, ultimately caused by our genes,…."


THE END