A computational framework for the global neuronal workspace theory

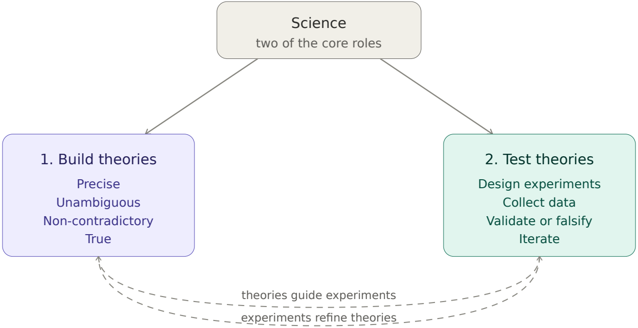
Gaspard Fougea, 3rd year PhD student at LMF under the direction of Alain Finkel and Stephane Le Roux
ENS Paris-Saclay

1. Motivations
2. Finite & timed automata
3. The GNW Theory and neuronal systems
4. Refinement & methodology
5. Theoretical usefulness of the modeling

Motivations

Motivations

Two roles of science



Motivations

- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)

Motivations

- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)

Motivations

- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)
- Difficulty to compare theories and to collaborate (verbal comparison) (Del Pin et al., 2021)

Motivations

- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)
- Difficulty to compare theories and to collaborate (verbal comparison) (Del Pin et al., 2021)
- Call for computational models to make theories more precise and falsifiable (Seth & Bayne, 2022; Kording et al., 2023)

Motivations

- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)
- Difficulty to compare theories and to collaborate (verbal comparison) (Del Pin et al., 2021)
- Call for computational models to make theories more precise and falsifiable (Seth & Bayne, 2022; Kording et al., 2023)
- In psychology, few existing methods to translate a verbal theory into a formal model:

Motivations

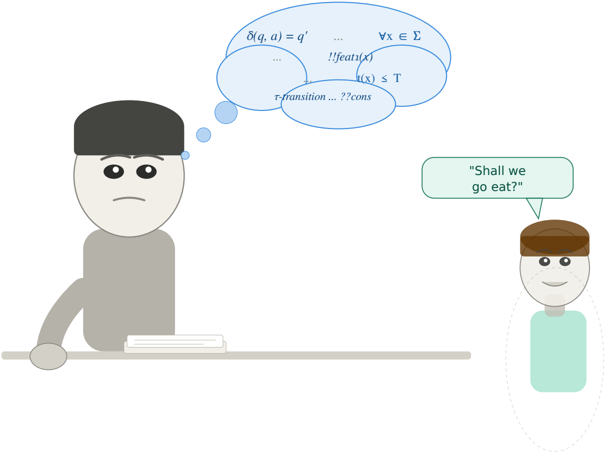
- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)
- Difficulty to compare theories and to collaborate (verbal comparison) (Del Pin et al., 2021)
- Call for computational models to make theories more precise and falsifiable (Seth & Bayne, 2022; Kording et al., 2023)
- In psychology, few existing methods to translate a verbal theory into a formal model:
  - Call to create frameworks and methodologies for theory specification and theory building (Glockner et al., 2025; ...)

Motivations

- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)
- Difficulty to compare theories and to collaborate (verbal comparison) (Del Pin et al., 2021)
- Call for computational models to make theories more precise and falsifiable (Seth & Bayne, 2022; Kording et al., 2023)
- In psychology, few existing methods to translate a verbal theory into a formal model:
  - Call to create frameworks and methodologies for theory specification and theory building (Glockner et al., 2025; ...)
  - Call to specify existing theories (Glockner et al., 2025; ...)

Motivations

Preconscious stimuli

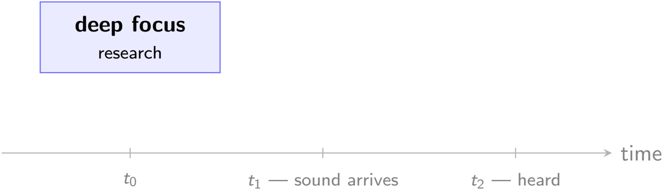


Motivations

Can we formalize the concept of "preconscious"?

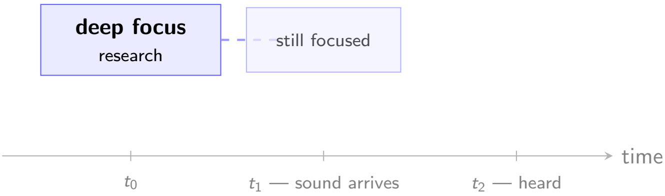
Motivations

Can we formalize the concept of “preconscious”?



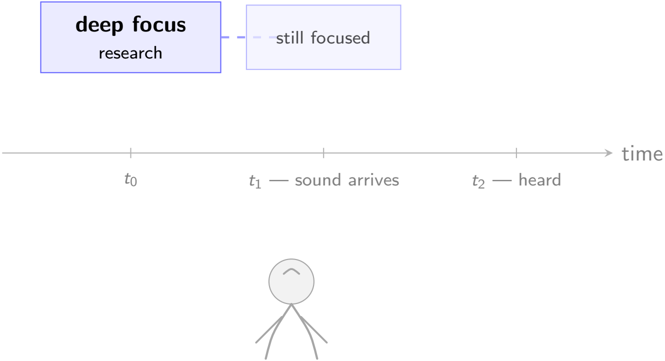
Motivations

Can we formalize the concept of “preconscious”?



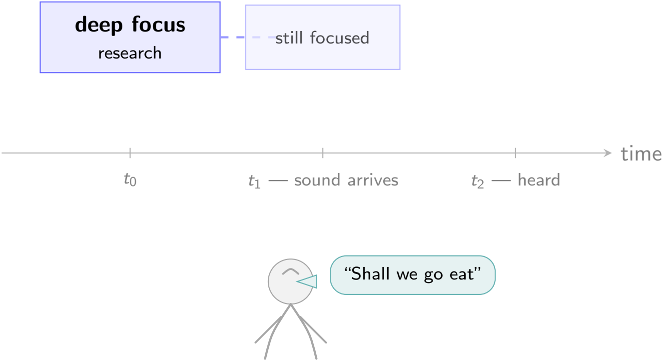
Motivations

Can we formalize the concept of "preconscious"?



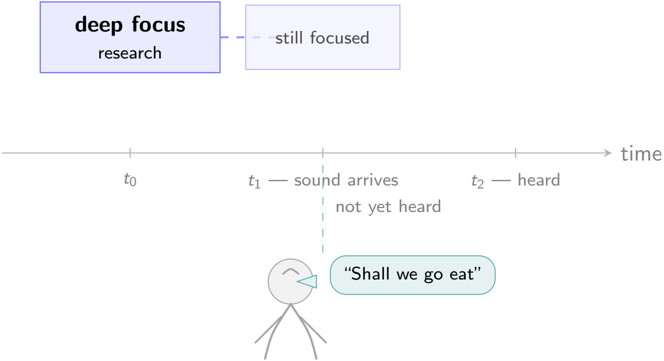
Motivations

Can we formalize the concept of "preconscious"?



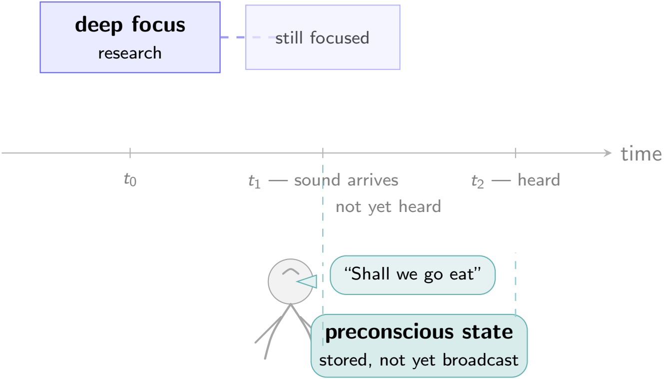
Motivations

Can we formalize the concept of "preconscious"?



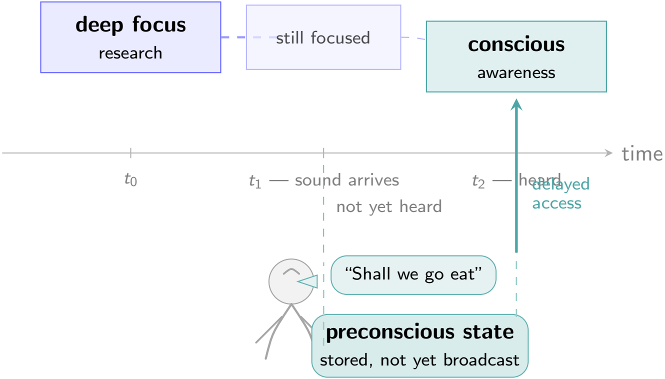
Motivations

Can we formalize the concept of “preconscious”?



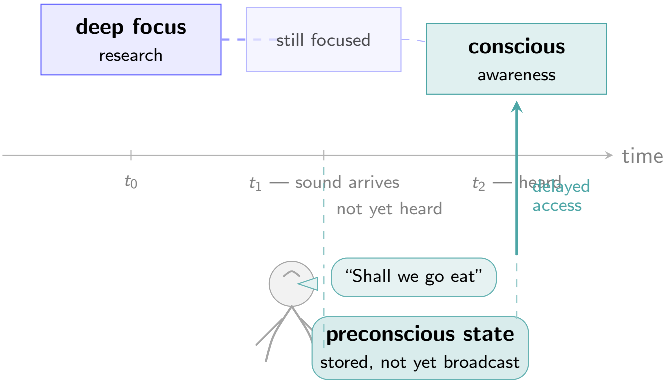
Motivations

Can we formalize the concept of “preconscious”?



Motivations

Can we formalize the concept of “preconscious”?



Can we formalize the concept of “preconscious”?

Motivations

Our Contribution: computational frameworks for theory building in psychology and consciousness

Frameworks from TCS:
Communicating (timed) automata, refinement, and model checking. Mathematical properties.

Methodology:
A methodology to translate a verbal theory into a formal model, then verify that the model is faithful to the verbal theory.

Application:
Formal model of the Global Neuronal Workspace Theory (Mashour et al., 2020) and of the Transactional Theory of Stress (Lazarus & Folkman, 1984).

Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,

Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,
2. Capability to handle large systems,

Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,
2. Capability to handle large systems,
3. Modularity (easy to modify, compose, and refine),

Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,
2. Capability to handle large systems,
3. Modularity (easy to modify, compose, and refine),
4. Having a formal dynamics, composition and refinement,

Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,
2. Capability to handle large systems,
3. Modularity (easy to modify, compose, and refine),
4. Having a formal dynamics, composition and refinement,
5. Possibility of step-by-step simulation,

Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,
2. Capability to handle large systems,
3. Modularity (easy to modify, compose, and refine),
4. Having a formal dynamics, composition and refinement,
5. Possibility of step-by-step simulation,
6. Formal verification of properties with the use of automatic tools,

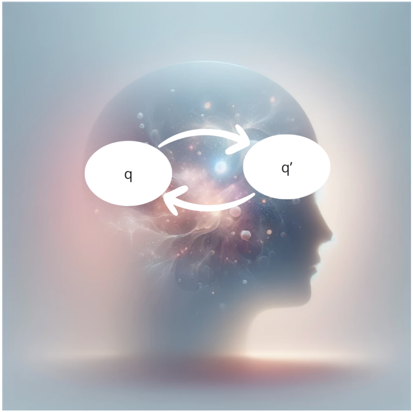
Motivations

Why this framework?

1. Openness to all psychological theories, both cognitive and behavioral,
2. Capability to handle large systems,
3. Modularity (easy to modify, compose, and refine),
4. Having a formal dynamics, composition and refinement,
5. Possibility of step-by-step simulation,
6. Formal verification of properties with the use of automatic tools,
7. Formal (and automatic) comparison of models, with automatic determination of compatibility between theories.

Motivations

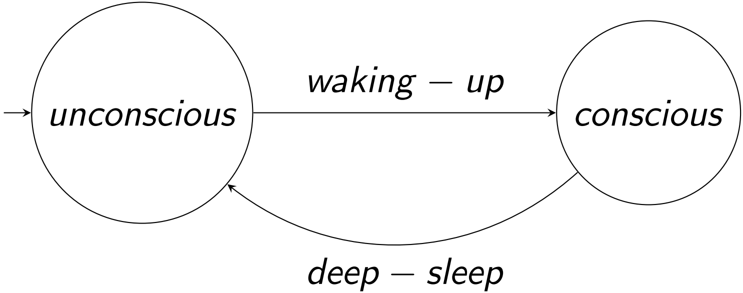
- We naturally identify states and transitions in (neuro)-cognitive theories



Motivations

- In (neuro)-cognitive theories, we naturally identify states and transitions
- Theoretical CS: powerful tools to model these theories. Finite automata & extensions!

Motivations



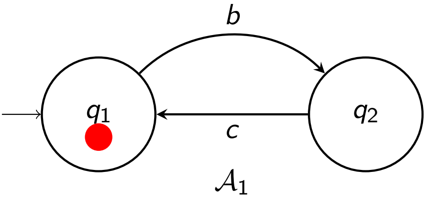
Basic understanding of consciousness

Finite & timed automata

Finite & timed automata

What is a finite automaton?

Example:

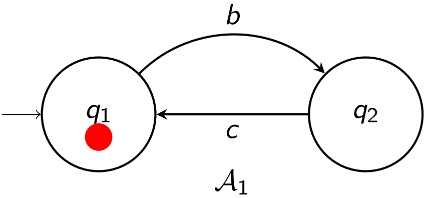


Finite & timed automata

What is a finite automaton?

Σ an alphabet
Q set of states
$\delta \subseteq Q \times \Sigma \times Q$ set of transitions
$I \subseteq Q$ initial states

Example:

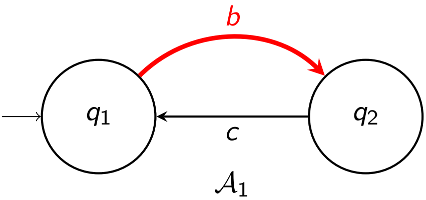


Finite & timed automata

What is a finite automaton?

Σ an alphabet
Q set of states
$\delta \subseteq Q \times \Sigma \times Q$ set of transitions
$I \subseteq Q$ initial states

Example:

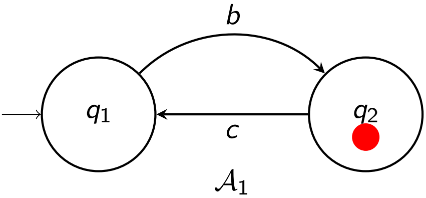


Finite & timed automata

What is a finite automaton?

Σ an alphabet
Q set of states
$\delta \subseteq Q \times \Sigma \times Q$ set of transitions
$I \subseteq Q$ initial states

Example:



Finite & timed automata

What is a timed automaton?

Σ an alphabet

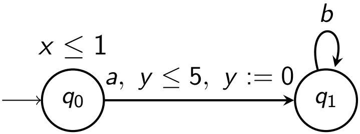
Q set of states

$\delta \subseteq Q \times \Sigma \times Q$ set of transitions

$I \subseteq Q$ initial states

A set of clocks {x, y}

State guards, transition guards, resets



$\mathcal{A}_{\textbf{timed}}$

Finite & timed automata

What is a timed automaton?

Σ an alphabet

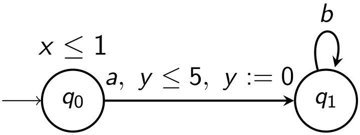
Q set of states

$\delta \subseteq Q \times \Sigma \times Q$ set of transitions

$I \subseteq Q$ initial states

A set of clocks {x, y}

State guards, transition guards, resets



$\mathcal{A}_{\textbf{timed}}$

Why timed automata for the GNW?

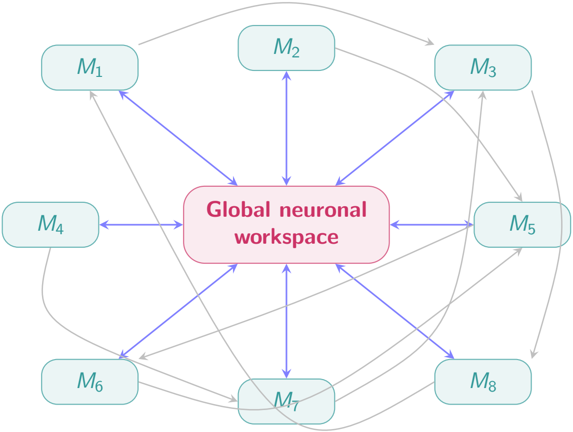
- Models time windows for preconscious stimuli, ignition, and activation of brain areas

The GNW Theory and neuronal systems

The GNW Theory and neuronal systems

Two computational spaces

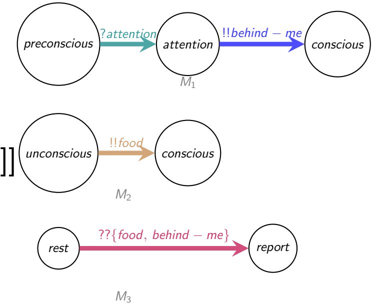
Modular circuits (modules) & The Global Neuronal Workspace

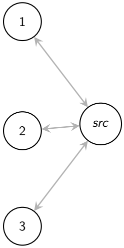


→ Bottom up and top-down transfers of information

The GNW Theory and neuronal systems

Neuronal systems.





GNW automaton

A neuronal system is $\mathcal{G}, \mathcal{M}_1, ..., \mathcal{M}_n$

$\mathcal{G}$: GNW automaton

$\mathcal{M}_i$: Modules

The GNW Theory and neuronal systems

Neuronal systems.

A neuronal system is
$\mathcal{G}, \mathcal{M}_1, ..., \mathcal{M}_n$
$\mathcal{G}$: GNW automaton
$\mathcal{M}_i$: Modules

The GNW Theory and neuronal systems

Neuronal systems.

A neuronal system is
$\mathcal{G}, \mathcal{M}_1, ..., \mathcal{M}_n$
$\mathcal{G}$: GNW automaton
$\mathcal{M}_i$: Modules

The states of $\mathcal{G}$ are $[n]$.
The set of states of the system is $Q = \bigcup_{i \in [n]} Q_i \cup [n]$

The GNW Theory and neuronal systems

Neuronal systems.

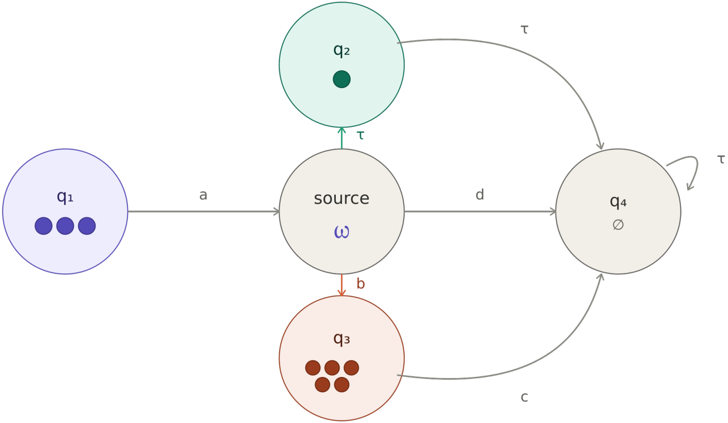
A neuronal system is
$\mathcal{G}, \mathcal{M}_1, ..., \mathcal{M}_n$
$\mathcal{G}$: GNW automaton
$\mathcal{M}_i$: Modules

The states of $\mathcal{G}$ are $[n]$.
The set of states of the system is $Q = \bigcup_{i \in [n]} Q_i \cup [n]$

A configuration of a neuronal system is a tuple (c, v), where $c : Q \to \mathbb{N}$ is a token function, and $v : X \to \mathbb{R}_+$ is a clock valuation.
States $\neq$ configurations

The GNW Theory and neuronal systems

$c : Q \to \mathbb{N}$ is a token function



The GNW Theory and neuronal systems

How the GNW mobilizes modules

The GNW can mobilize certain brain areas to broadcast their content while suppressing others. (Dehaene & Naccache, 2001)

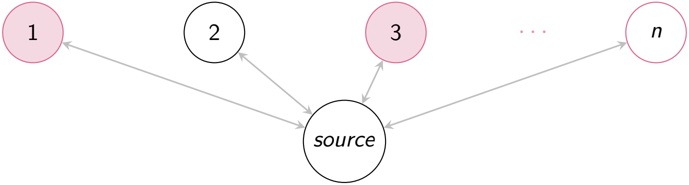
The GNW Theory and neuronal systems

How the GNW mobilizes modules

The GNW can mobilize certain brain areas to broadcast their content while suppressing others. (Dehaene & Naccache, 2001)

In a configuration (c, v), a module $\mathcal{M}_i$ is:

- **mobilized** if $c(i) > 0$
- **suppressed** if $c(i) = 0$



GNW automaton

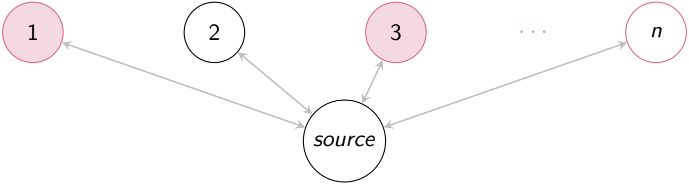
The GNW Theory and neuronal systems

How the GNW mobilizes modules

The GNW can mobilize certain brain areas to broadcast their content while suppressing others. (Dehaene & Naccache, 2001)

In a configuration (c, v), a module $\mathcal{M}_i$ is:

- **mobilized** if $c(i) > 0$
- **suppressed** if $c(i) = 0$



GNW automaton

Mobilized modules are the ones that contribute to a conscious experience.

The GNW Theory and neuronal systems

Conscious processing (GNW Theory)

"A perceptual content is conscious when it is globally broadcast to many other local brain areas, making it available for verbal report and motor control." (Mashour et al., 2020)

The GNW Theory and neuronal systems

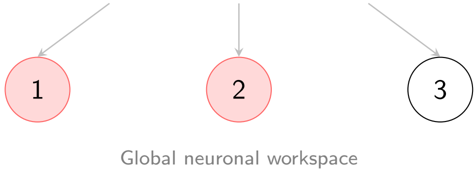
Conscious processing (GNW Theory)

"A perceptual content is conscious when it is globally broadcast to many other local brain areas, making it available for verbal report and motor control." (Mashour et al., 2020)

"The GNW possesses the ability to receive bottom-up information from and transmit top-down information to any of the various processors." (Mashour et al., 2020)

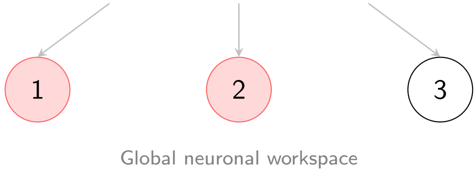
The GNW Theory and neuronal systems

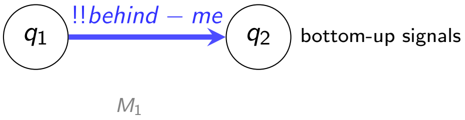
Conscious broadcast (neuronal systems)



The GNW Theory and neuronal systems

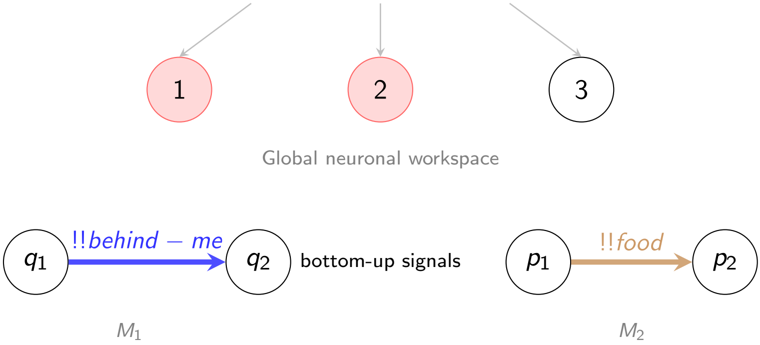
Conscious broadcast (neuronal systems)





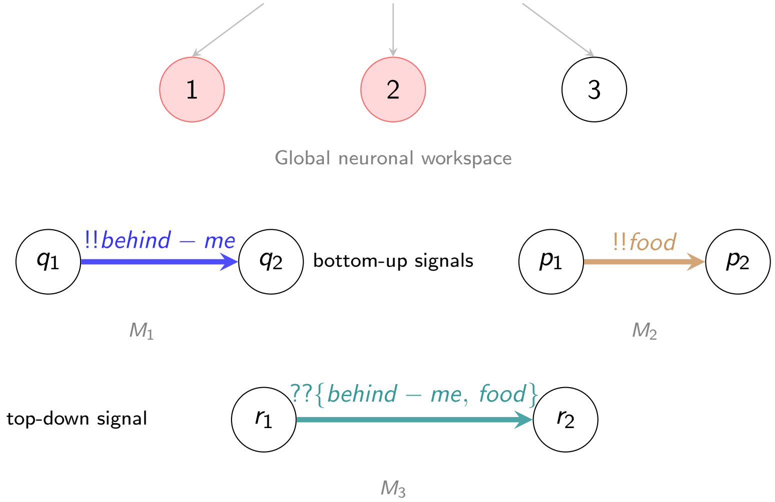
The GNW Theory and neuronal systems

Conscious broadcast (neuronal systems)



The GNW Theory and neuronal systems

Conscious broadcast (neuronal systems)



conscious broadcast integrates all features into a

The GNW Theory and neuronal systems

Local communication between modules

An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)

The GNW Theory and neuronal systems

Local communication between modules

An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)

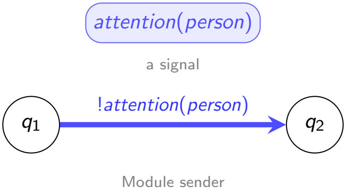
$attention(person)$

a signal

The GNW Theory and neuronal systems

Local communication between modules

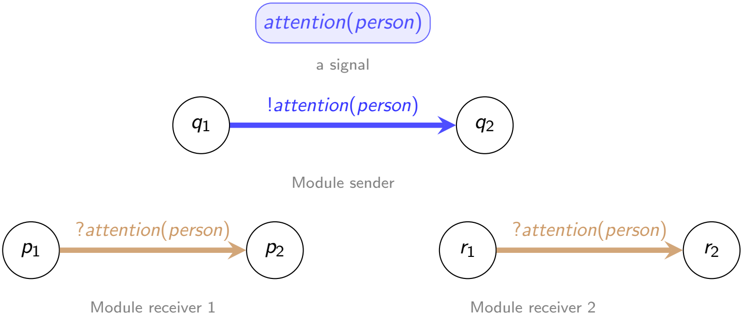
An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)



The GNW Theory and neuronal systems

Local communication between modules

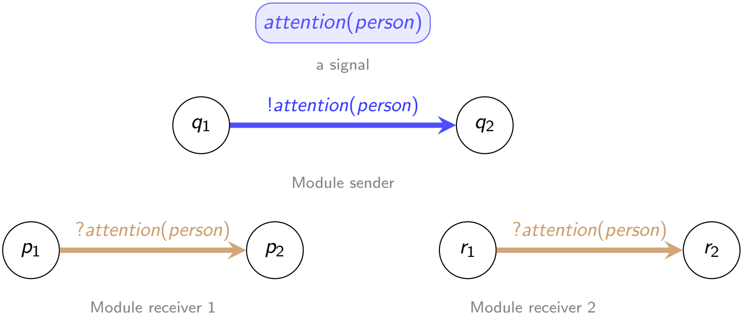
An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)



The GNW Theory and neuronal systems

Local communication between modules

An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)



when !*attention*(*person*) fires, **all** automata having a ?*attention*(*person*) transitions fire simultaneously

The GNW Theory and neuronal systems

What can we model with neuronal systems?

Any theory that has conscious and/or unconscious components.

Decision making theories, theories of emotion, reasoning, perception, memory, etc.

Refinement & methodology

Refinement & methodology

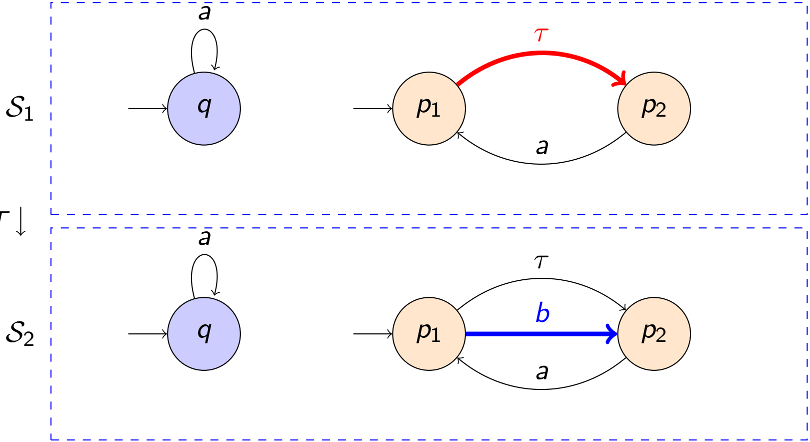
- Nine refinement operations to develop a theory step by step
- Start from a simple (possibly trivial) model and refine iteratively

Operations:

- Adding / removing a transition
- Splitting a state
- Adding a state guard, transition guard
- Adding a basic automaton

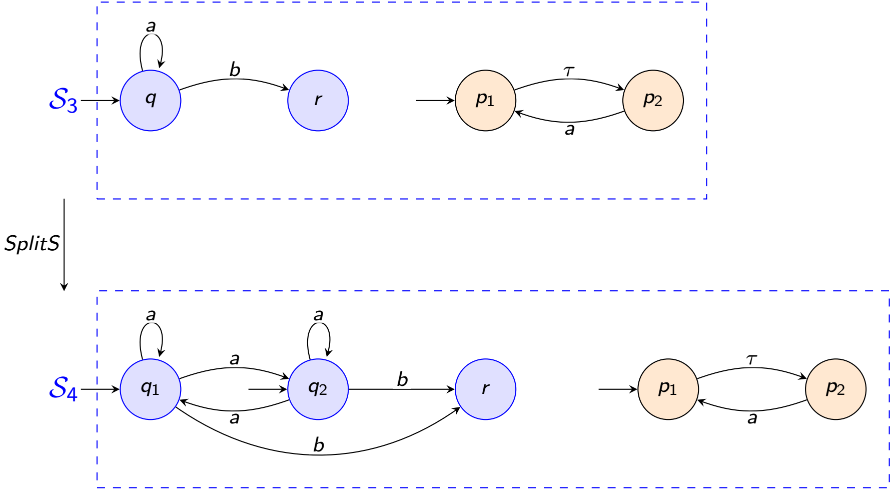
Refinement & methodology

Adding a transition



Refinement & methodology

Splitting a state



Refinement & methodology

General methodology

Empty system

Refinement & methodology

General methodology

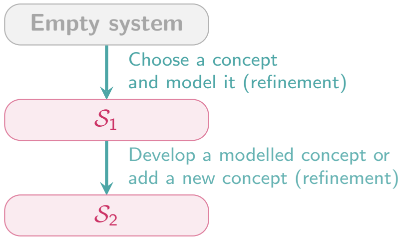
Empty system

$\downarrow$ Choose a concept and model it (refinement)

$\mathcal{S}_1$

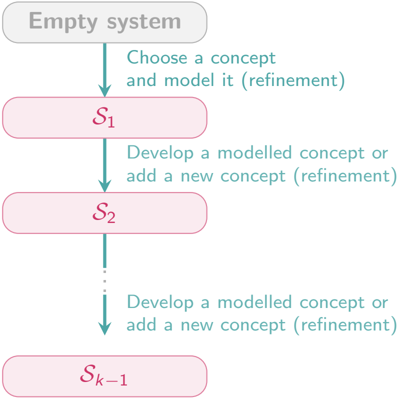
Refinement & methodology

General methodology



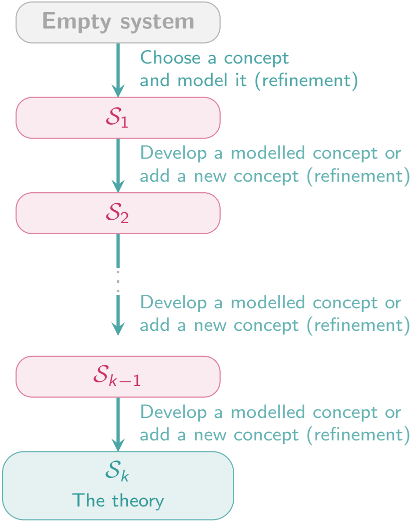
Refinement & methodology

General methodology



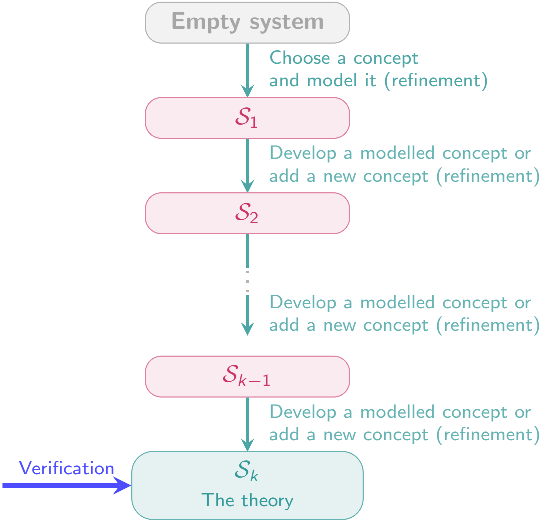
Refinement & methodology

General methodology



Refinement & methodology

General methodology



Refinement & methodology

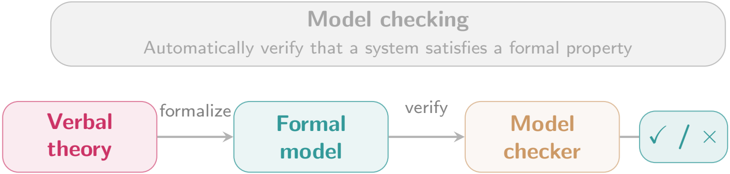
The use of verification

Model checking

Automatically verify that a system satisfies a formal property

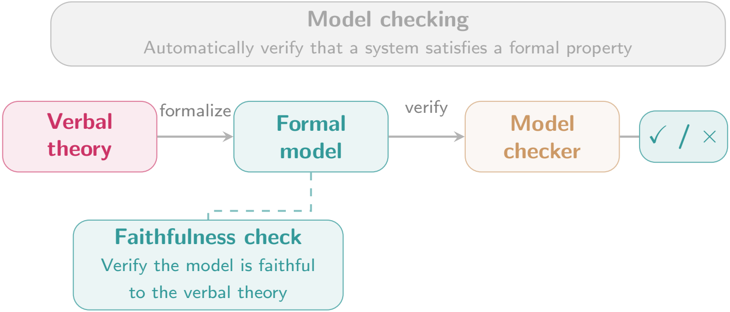
Refinement & methodology

The use of verification



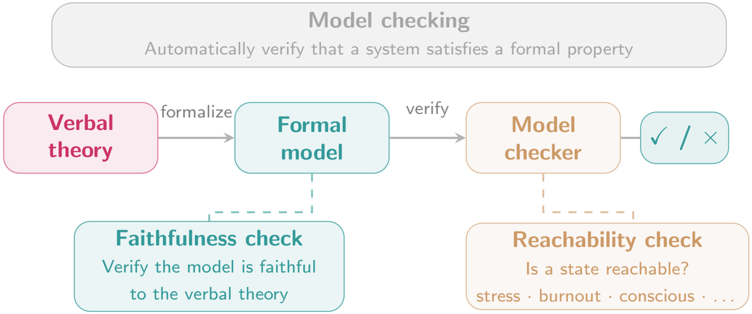
Refinement & methodology

The use of verification



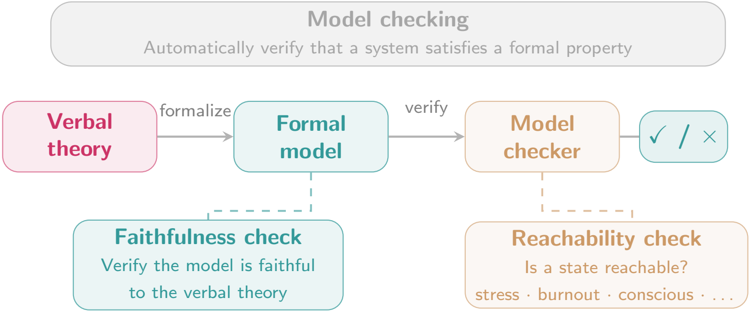
Refinement & methodology

The use of verification



Refinement & methodology

The use of verification



psychological model checking

Refinement & methodology

Modeling choices arise when...

Refinement & methodology

Modeling choices arise when. . .

Refinement & methodology

Modeling choices arise when. . .

Type 1
Multiple models equivalently fit the text

Refinement & methodology

Modeling choices arise when. . .

Type 1
Multiple models equivalently fit the text

Type 2
Theory gives no info on inevitable steps

Refinement & methodology

Modeling choices arise when. . .

Type 1
Multiple models equivalently fit the text

Type 2
Theory gives no info on inevitable steps

Type 3
Two or more meanings can be derived

Refinement & methodology

Modeling choices arise when. . .

Type 1
Multiple models equivalently fit the text

Type 2
Theory gives no info on inevitable steps

Type 3
Two or more meanings can be derived

Theoretical usefulness
Identified underspecifications · Ambiguities · research questions

Refinement & methodology

Modeling choices arise when. . .

Type 1
Multiple models equivalently fit the text

Type 2
Theory gives no info on inevitable steps

Type 3
Two or more meanings can be derived

Theoretical usefulness
Identified underspecifications · Ambiguities · research questions

Principles for modeling choices

Equivalent options?
Choose the easier to understand

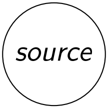
Local concept? → model as a state
Shared concept? → model as a letter

No set rules — these are guiding principles

Refinement & methodology

Let us use this methodology!

We begin with the empty system $(\varnothing, \varnothing, \varnothing, \mathcal{G}_0)$



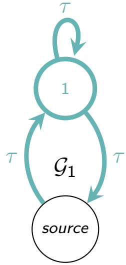
$\mathcal{G}_0$

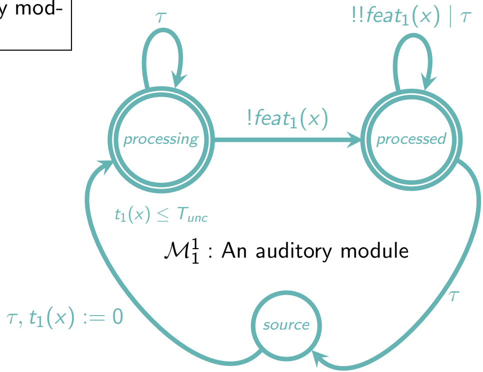
Refinement & methodology

An auditory module

T_{unc} : time for the unconscious processing of a stimulus to be processed by the auditory module

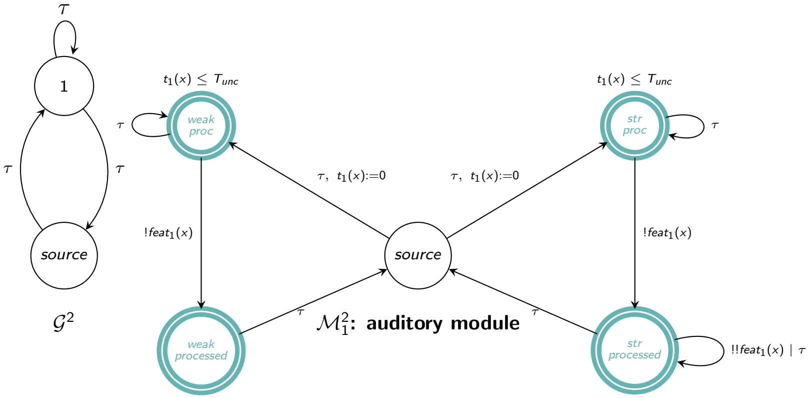
$\boxed{Param = Stim}$





Refinement & methodology

Strong and subliminal stimuli



$\mathcal{M}_1^2$: **auditory module**

Refinement & methodology

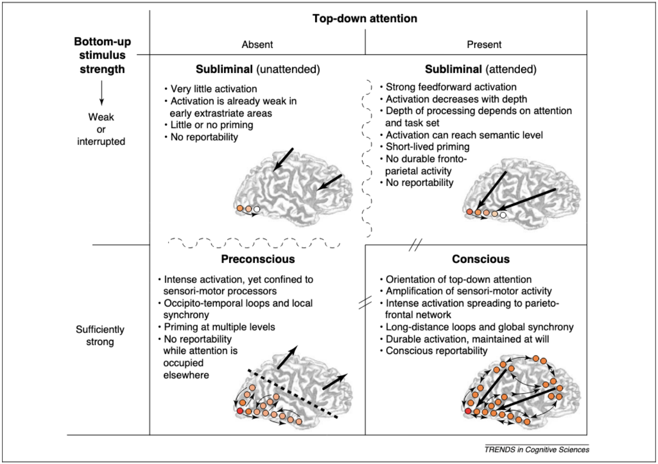
Attention and preconscious stimuli

Attention selects signals and amplifies them (Mashour et al., 2020)
Preconscious stimuli: strong signal but no attention. (Dehaene et al., 2006)

Refinement & methodology

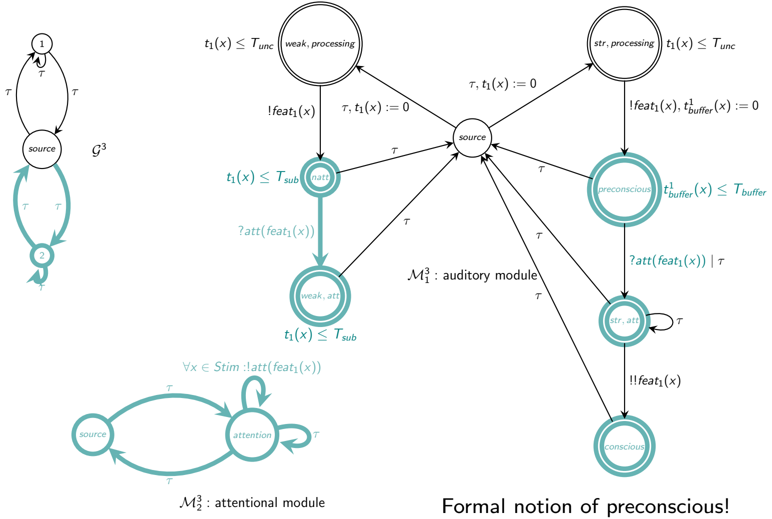
Attention and preconscious stimuli

(Dehaene et al., 2006)



Refinement & methodology

Attention and preconscious stimuli

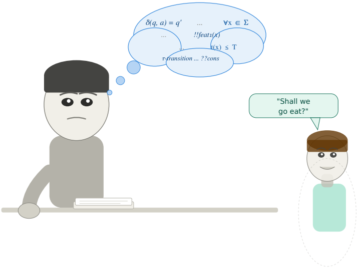


Formal notion of preconscious!
(and conscious)

Refinement & methodology

A run of the system

$$(\omega \cdot \mathit{source}) \xrightarrow{\tau} (1 \cdot (x, \mathit{str}, \mathit{processing})) \xrightarrow{!\mathit{feat}_1(x)}$$
$$(1 \cdot (x, \mathit{preconscious})) \xrightarrow{300} (1 \cdot (x, \mathit{preconscious})) \xrightarrow{!\mathit{att}(\mathit{feat}_1(x))}$$
$$(1 \cdot (x, \mathit{str}, \mathit{att})) \xrightarrow{!!\{\mathit{feat}_1(x)\}} (1 \cdot (x, \mathit{conscious}))$$



Refinement & methodology

Discussion

Not a representation of what happens at the level of neurons.
However, we could refine to the level of neurons.

Theoretical usefulness of the modeling

Theoretical usefulness of the modeling

Formalization of a theory

three types of results emerge

Theoretical usefulness of the modeling

Formalization of a theory
three types of results emerge

Ambiguities
& inconsistencies
in the verbal theory

Theoretical usefulness of the modeling

Formalization of a theory
three types of results emerge

Ambiguities
& inconsistencies
in the verbal theory

Under-specifications
"holes" in the
dynamics

Theoretical usefulness of the modeling

Formalization of a theory
three types of results emerge

- **Ambiguities**
  & inconsistencies
  in the verbal theory
- **Under-specifications**
  "holes" in the
  dynamics
- **Logical properties**
  implied by the theory
  not in the text

Theoretical usefulness of the modeling

Formalization of a theory
three types of results emerge

Ambiguities
& inconsistencies
in the verbal theory

Under-specifications
"holes" in the
dynamics

Logical properties
implied by the theory
not in the text

Research questions & theory refinement
all three types prompt further scientific inquiry

Ambiguities

- When is a content conscious? During ignition? When it is broadcast? When it is received? For how long?
- Ambiguity about the 'conscious threshold'. Are attention + stimulus strength sufficient?

Ambiguities

- When is a content conscious? During ignition? When it is broadcast? When it is received? For how long?
- Ambiguity about the 'conscious threshold'. Are attention + stimulus strength sufficient?

Under-specifications

- When does a content exists consciousness?
- How do we transit from one conscious content to another?
- What is the time between ignition and broadcast?
- Is there a minimum time for a content to be conscious?

Theoretical usefulness of the modeling

Main result: double ignition

Two sentences from (Mashour et al., 2020).

- *"the final result of attentional selection enters consciousness, where it can use the GNW to activate all features that belong to the same object even if they are represented in different cortical areas (Roelfsema and Houtkamp, 2011)"* (Mashour et al. 2020)
- *"Ignition is characterized by the sudden, coherent, and exclusive activation of a subset of workspace neurons coding for the current conscious content, with the remainder of the workspace neurons being inhibited."* (Mashour et al. 2020)

Theoretical usefulness of the modeling

Main result: double ignition

Two sentences from (Mashour et al., 2020).

- *"the final result of attentional selection enters consciousness, where it can use the GNW to activate all features that belong to the same object even if they are represented in different cortical areas (Roelfsema and Houtkamp, 2011)"* (Mashour et al. 2020)
- *"Ignition is characterized by the sudden, coherent, and exclusive activation of a subset of workspace neurons coding for the current conscious content, with the remainder of the workspace neurons being inhibited."* (Mashour et al. 2020)

Together imply:

- There are two ignitions when attention is brought to a particular stimulus.
- **"Preconscious"** applies to *features*, not to stimuli
- New properties to test to validate/falsify the theory!

Theoretical usefulness of the modeling

Theorem 1. (Untimed) neuronal systems are equivalent to counter automata

Counter automata are very analysed, verifiable models.
Hence, reachability is decidable for untimed neuronal systems.

Theoretical usefulness of the modeling

Proposition. The refinement operations allow us to create any system from a trivial system.

Theoretical usefulness of the modeling

Proposition. The refinement operations allow us to create any system from a trivial system.

Theorem 2. There exists a (non-trivial) relation $\mathcal{R}$ such that if $\mathcal{S}'$ is a refinement of $\mathcal{S}$, then for all runs w' of $\mathcal{S}'$, there exists a run w of $\mathcal{S}$ such that $w \mathrel{\mathcal{R}} w'$.

Theoretical usefulness of the modeling

Proposition. The refinement operations allow us to create any system from a trivial system.

Theorem 2. There exists a (non-trivial) relation $\mathcal{R}$ such that if $\mathcal{S}'$ is a refinement of $\mathcal{S}$, then for all runs w' of $\mathcal{S}'$, there exists a run w of $\mathcal{S}$ such that $w \; \mathcal{R} \; w'$.

Interpretation: Refinement accounts to constraining. When refining, new behaviors do not appear.

Conclusion & Perspectives

Results:

- A formal framework (neuronal systems + refinement) to build theories grounded in the GNW Theory
- A methodology to translate such verbal theories into formal models
- A formal model of the GNW Theory
- The detection of a theoretical property, detected ambiguities and lack of specification as well as prompted research questions.
- Mathematical results to analyse neuronal systems and for refinement

Conclusion & Perspectives

Perspectives

- Create a merging operator to combine different theories of consciousness
- Formally compare GNW with other theories (IIT, HOT, ...)
- Create a modular, composite theory by merging compatible theories
- Develop a graphical interface for modeling and model checking (UPPAAL / TCHECKER)

Conclusion & Perspectives

Events.

1st workshop on Computational Psychology

June 10th at ENS Paris-Saclay

Conclusion & Perspectives

Events.

1st workshop on Computational Psychology
June 10th at ENS Paris-Saclay

Training session: building formal models with the framework
June 9th at ENS Paris-Saclay
June 8th (zoom)

Conclusion & Perspectives

[1] Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace Theory. *Neuron*, 105(5), 776–798.
[2] Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.
[3] Meehl, P. E. (1978). Theoretical risks and tabular asterisks. *Journal of Consulting and Clinical Psychology*, 46(4), 806.
[4] Robinaugh, D. J., et al. (2021). Invisible hands and fine calipers: A call to use formal theory as a toolkit for theory construction.
[5] Alur, R., & Dill, D. L. (1994). A theory of timed automata. *Theoretical Computer Science*, 126(2), 183–235.
[6] Horn, F., & Sangnier, A. (2020). Deciding the existence of minority terms. *LICS*.