Modeling the global neuronal workspace theory with communicating timed automata

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

Motivations

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

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

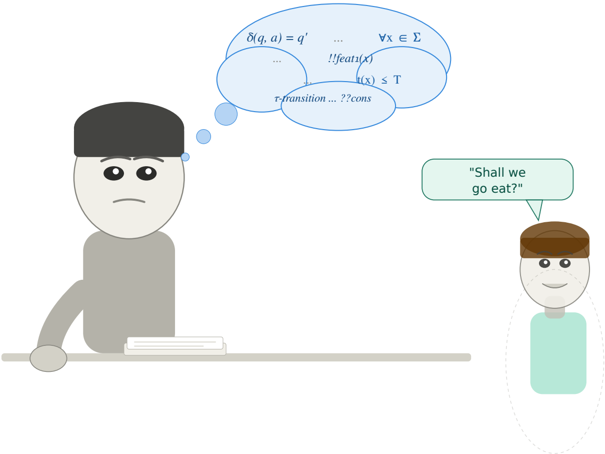
- Theories of consciousness are often underspecified (verbal theories) (Seth & Bayne, 2022; Butlin et al., 2023)
- Difficulty to validate/falsify imprecise theories (Robinaugh et al., 2021)
- Difficulty to compare theories and to collaborate (verbal comparison) (Del Pin et al., 2021)
- Call for computational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

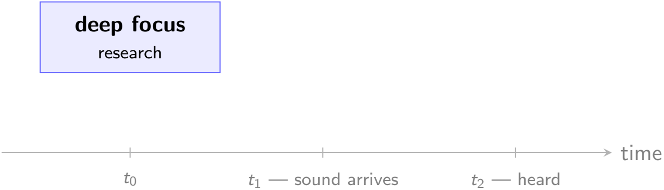


Motivations

Can we formalize the concept of “preconscious”?

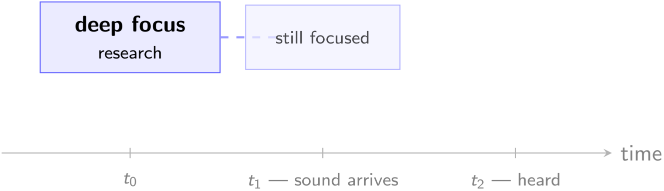
Motivations

Can we formalize the concept of "preconscious"?



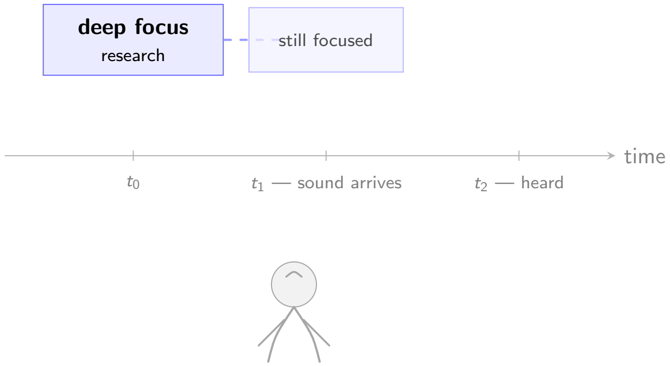
Motivations

Can we formalize the concept of "preconscious"?



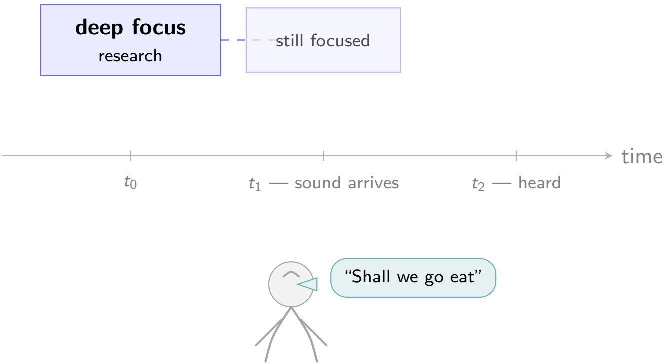
Motivations

Can we formalize the concept of “preconscious”?



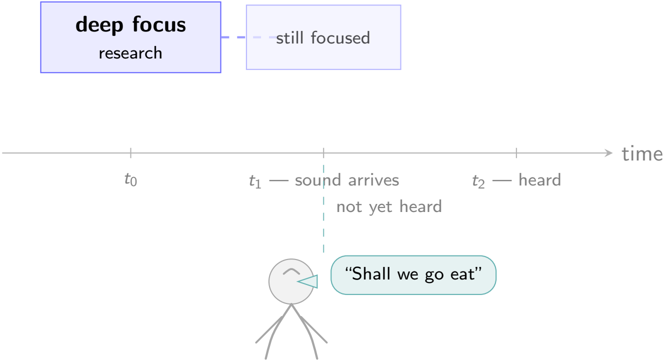
Motivations

Can we formalize the concept of “preconscious”?



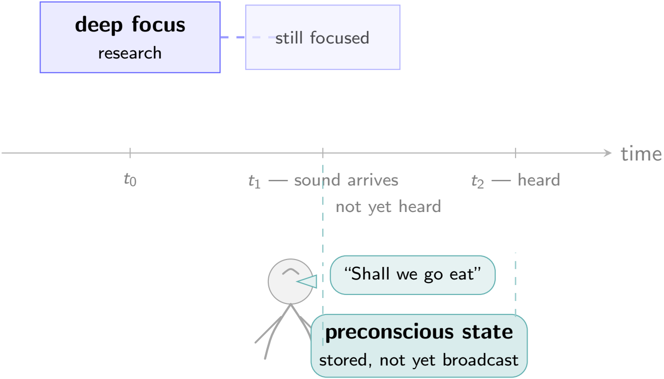
Motivations

Can we formalize the concept of "preconscious"?



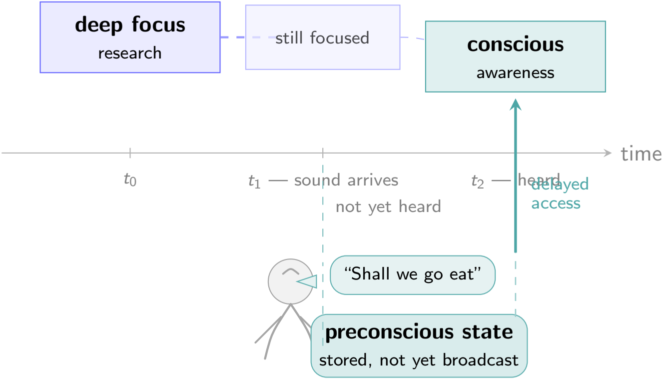
Motivations

Can we formalize the concept of “preconscious”?



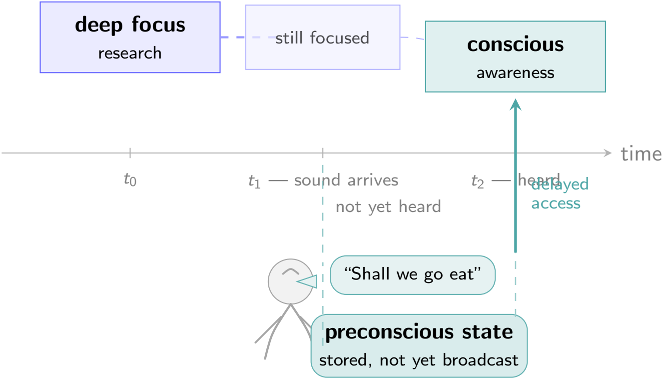
Motivations

Can we formalize the concept of “preconscious”?



Motivations

Can we formalize the concept of "preconscious"?



Can we formalize the concept of "preconscious"?

Motivations

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,

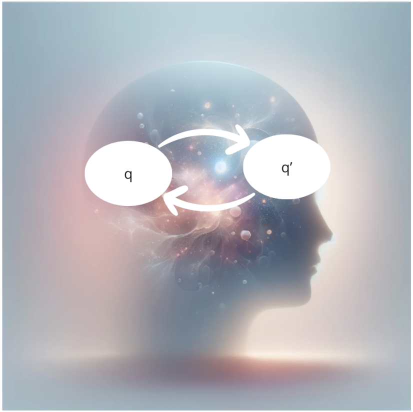
Motivations

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

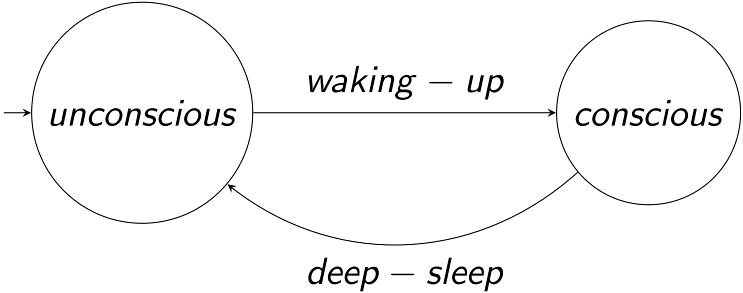
- We naturally identify states and transitions in (neuro)-cognitive theories



Motivations

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

Motivations



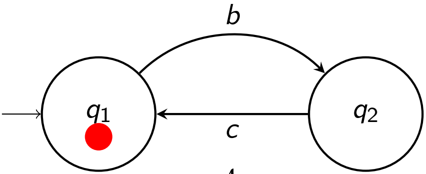
Basic understanding of consciousness

Finite & timed automata

Finite & timed automata

What is a finite automaton?

Example:



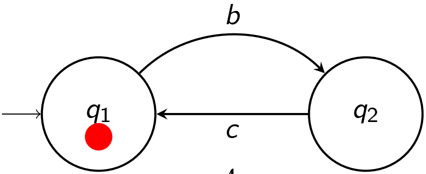
$\mathcal{A}_1$

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:



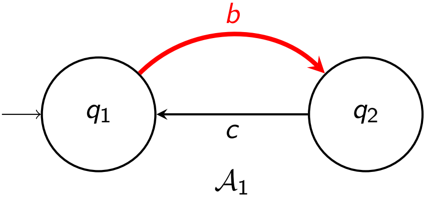
$\mathcal{A}_1$

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:

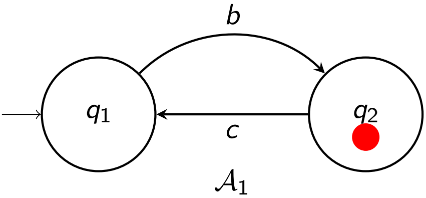


Finite & timed automata

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

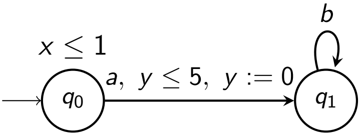
Q set of states

$\delta \subseteq Q \times \Sigma \times Q$ 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

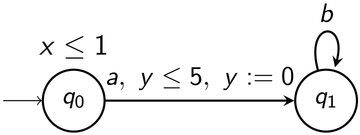
Q set of states

$\delta \subseteq Q \times \Sigma \times Q$ 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?

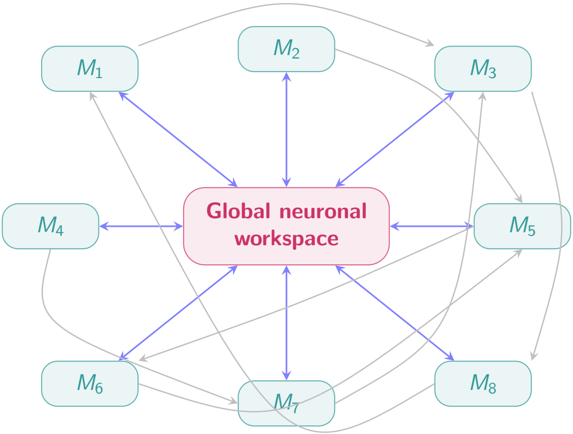
- Models time windows for preconscious stimuli, ignition, and activation of brain areas

The GNW Theory and neuronal systems

The GNW Theory and neuronal systems

Two computational spaces

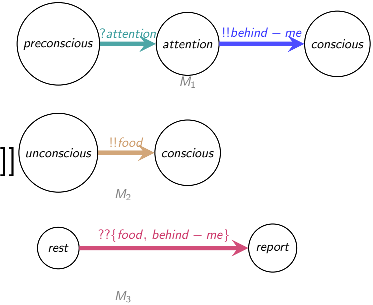
Modular circuits (modules) & The Global Neuronal Workspace

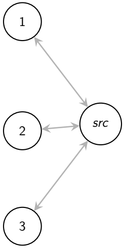


→ Bottom up and top-down transfers of information

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.

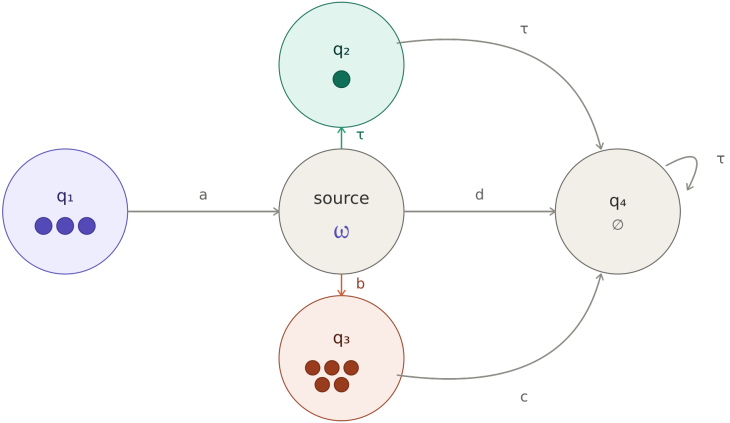
A neuronal system is
$\mathcal{G}, \mathcal{M}_1, ..., \mathcal{M}_n$
$\mathcal{G}$: GNW automaton
$\mathcal{M}_i$: Modules

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)

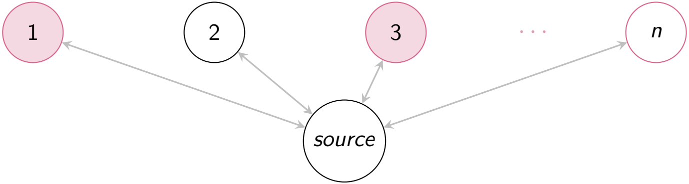
The GNW Theory and neuronal systems

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

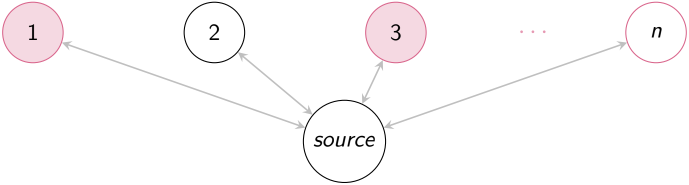
The GNW Theory and neuronal systems

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

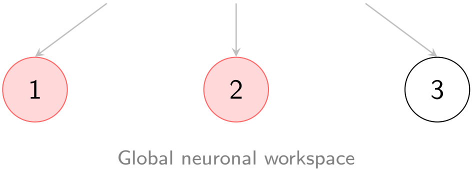
Conscious processing (GNW Theory)

"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)

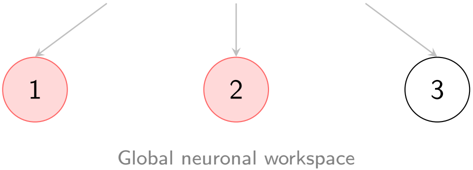
The GNW Theory and neuronal systems

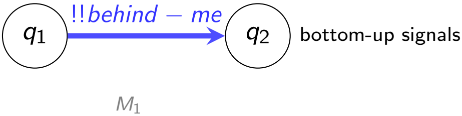
Conscious broadcast (neuronal systems)



The GNW Theory and neuronal systems

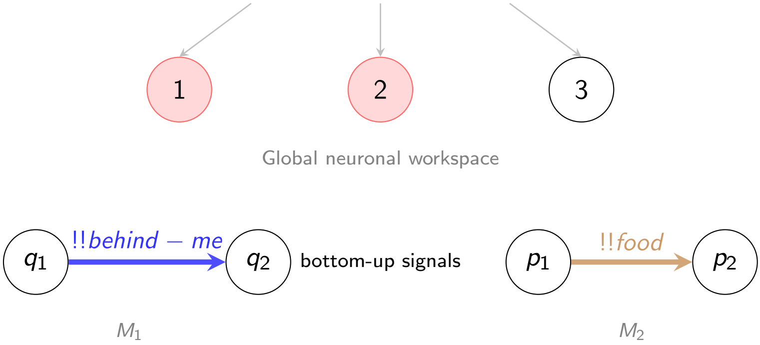
Conscious broadcast (neuronal systems)





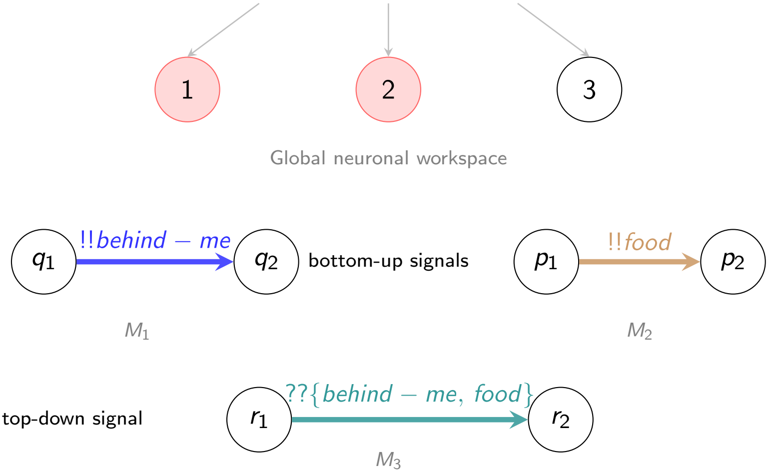
The GNW Theory and neuronal systems

Conscious broadcast (neuronal systems)



The GNW Theory and neuronal systems

Conscious broadcast (neuronal systems)



conscious broadcast integrates all features into a "coherent, multifeature, and multimodal" object

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)

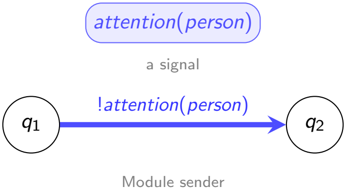
$\boxed{attention(person)}$

a signal

The GNW Theory and neuronal systems

Local communication between modules

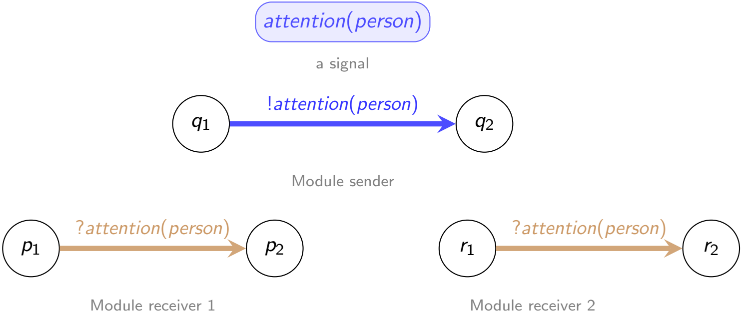
An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)



The GNW Theory and neuronal systems

Local communication between modules

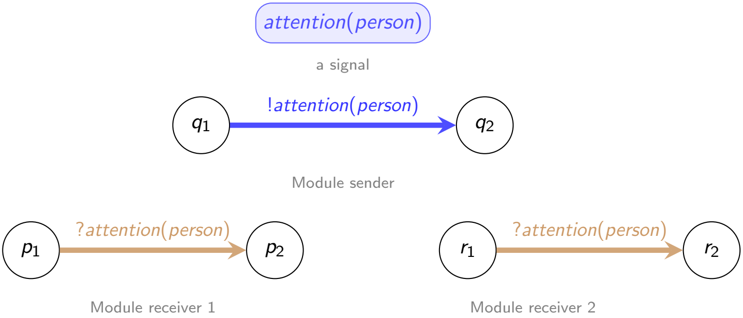
An unconscious operation can occur if the brain areas / modules involved are connected. (Dehaene & Naccache, 2001)



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

Main result

Theorem.

- For every neuronal system $\mathcal{S}$, there exists a Petri net $\mathcal{P}_{\mathcal{S}}$ that is in a transitive simulation with the untimed transition system of $\mathcal{S}$;
- For every Petri net $\mathcal{P}$ with an empty initial marking and arcs bounded by one, there exists a neuronal system $\mathcal{S}_{\mathcal{P}}$ such that its untimed transition system is in bisimulation with $\mathcal{P}$.

$\rightarrow$ Reachability and coverability problems are decidable for untimed Neuronal systems.

The GNW Theory and neuronal systems

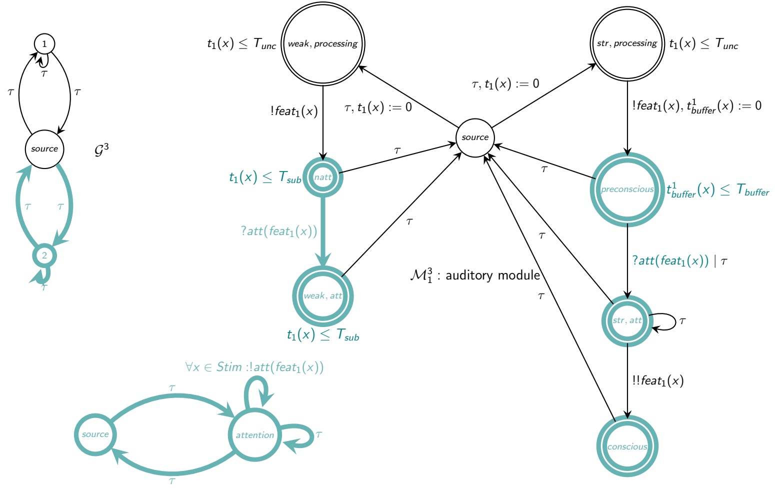
What can we model with neuronal systems?

Any theory that has conscious and/or unconscious components.

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

The GNW Theory and neuronal systems

A neuronal system



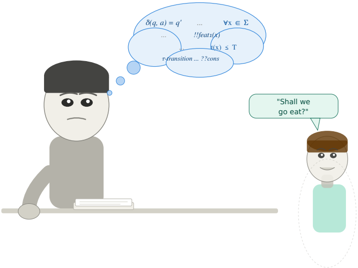
$\mathcal{M}^3_2$: attentional module

Formal notion of preconscious!
(and conscious)

The GNW Theory and neuronal systems

A run of the system

$$(\omega \cdot source) \xrightarrow{\tau} (1 \cdot (x, str, processing)) \xrightarrow{!feat_1(x)}$$
$$(1 \cdot (x, preconscious)) \xrightarrow{300} (1 \cdot (x, preconscious)) \xrightarrow{!att(feat_1(x))}$$
$$(1 \cdot (x, str, att)) \xrightarrow{!!\{feat_1(x)\}} (1 \cdot (x, conscious))$$



The GNW Theory and neuronal systems

Discussion

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

The GNW Theory and neuronal systems

Discussion

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

We do not claim that neuronal systems are conscious!

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

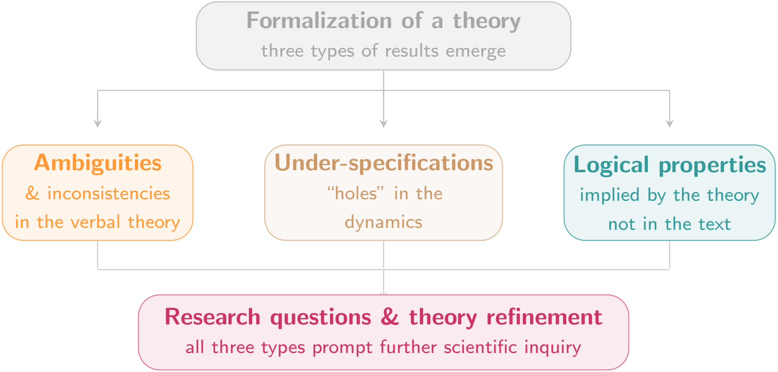
Under-specifications
"holes" in the
dynamics

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



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!

Conclusion & Perspectives

Results:

- A formal framework (neuronal systems) to build theories grounded in the GNW Theory
- 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 result to analyse neuronal systems

Conclusion & Perspectives

Perspectives

- Creation of refinement operations for theory-building + mathematical properties
- 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)
- Training researchers in our theory-building frameworks (next one on July 28th)

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*.