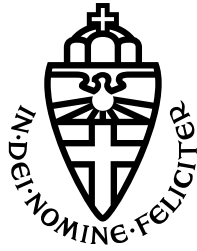


RADBOUD UNIVERSITY NIJMEGEN



FACULTY OF SOCIAL SCIENCES

Flexible serial processing in networks of spiking neurons

FROM ACCESS CONSCIOUSNESS TO FLEXIBLE BEHAVIOUR

MASTER'S THESIS IN ARTIFICIAL INTELLIGENCE

Author:

Hugo CHATEAU-LAURENT

Student number:

1023970

Internal supervisor:

Serge THILL

Radboud University

External supervisor:

Chris ELIASMITH

University of Waterloo

Second reader:

Johan KWISTHOUT

Radboud University

August 2020

Abstract

The ability to flexibly route information between otherwise separate brain regions is crucial to perform newly instructed tasks in which each step takes as input the result of the preceding step. How the brain implements that in parallel substrate is not well understood. Psychological studies suggest that consciousness plays a key role, as single steps can be performed when the input stimulus is subliminal, but not multiple ones. This is in line with the global neuronal workspace theory which suggests that information is consciously perceived when granted access to a broad fronto-parietal network and can be broadcast to virtually any processor in the brain.

In this project, I propose the first spiking neural model based on this theory to perform serial tasks composed of multiple steps. The Neural Engineering Framework and the Semantic Pointer Architecture are used to construct the model consisting of parallel processing networks, a global workspace, working memory components and gating networks. The resulting neural network is consistent with global neuronal workspace theory at both cognitive and biological levels, as it accounts for conscious competition, the associated ignition indexed by P3 waves, and selective broadcast. Interestingly, it also addresses the unsolved issue of diachronic and synchronic integration in the process of forming coalitions for conscious access. Simulations suggest that the model is capable of performing many steps in a robust manner. Most importantly, it reproduces the same patterns of response time as humans. Hence, the model provides a good framework for explaining the role of consciousness in serial processing.

Contents

1	Introduction	4
1.1	Goals	4
1.2	Organisation	4
2	Background	5
2.1	Consciousness	5
2.1.1	Historical background	5
2.1.2	Minimal contrast paradigms	6
2.1.3	Global workspace theory	6
2.1.4	Neural implementations	9
2.1.5	Global neuronal workspace simulations	12
2.2	Flexible sequential processing	13
2.2.1	Experiment 1: chronometric exploration	14
2.2.2	Experiment 3: priming	16
2.2.3	Experiment 4: backward masking	17
3	Methods	20
3.1	Neural Engineering Framework	20
3.1.1	Representation	20
3.1.2	Computation	21
3.1.3	Dynamics	21
3.2	Semantic Pointer Architecture	22
3.3	Parameter optimisation	23
4	Model	26
4.1	Processors	27
4.1.1	Sensory and motor	27
4.1.2	Hetero-associative memory	27
4.1.3	Accumulation-to-threshold	28
4.2	Global workspace	29
4.3	Processors-workspace interactions	31

4.4	Router	32
4.4.1	Working memory	32
4.4.2	Production rules	33
5	Results	36
5.1	Model behaviour	36
5.2	Empirical fitness	38
5.2.1	Parameters	38
5.2.2	Optimisation	39
5.3	Performance on long tasks	42
6	Discussion	45

Acknowledgements

I wish to express my sincere appreciation to my supervisors. Serge Thill has been of great support all along my Master's program and has always been available to clarify my doubts and fulfil my scientific curiosity. Chris Eliasmith has paved the way for the neural modelling of higher-level cognition, and hence, for my research project. He warmly welcomed me in Canada and provided very useful guidelines. Without their help, the goal of this project would not have been realised.

To the team at the Computational Neuroscience Research Group, my deepest thanks for your welcome, most especially to Terrence Stewart for taking the time to answer my questions, and Narsimha Chilkuri who helped me with the mathematical formalism and became a true friend.

The financial contribution of Radboud University was truly appreciated for helping me travel to Canada.

I am grateful to my mother and sibling Anne and Sarah for their feedback and continued love. Thank you to Lydia, for all her love and support. I am also deeply grateful to Daphne, Masha and Victor for being thought-provoking, and my other friends who have supported me along the way.

Finally, I wish to dedicate this thesis to my sorely missed friend Léo, with whom I had the most profound scientific conversations and first got enthusiastic about the study of consciousness.

Chapter 1

Introduction

1.1 Goals

In the last three decades, tremendous progress has been made at understanding consciousness and its neural basis. A theory rated as the most promising by experts (Michel et al., 2018) has emerged: the global workspace theory. However the following question has not yet been addressed: “How can a neural model with a global workspace architecture perform multi-step tasks and account for human behaviour in these tasks?”. Hence, this is the question I will tackle here.

1.2 Organisation

In Background (chapter 2), I describe the study of consciousness and human serial processing. In chapter 3, I explain the methods for creating and optimising the model, which is described in chapter 4. The behaviour of the model is described in Results (chapter 5) and discussed in chapter 6.

Chapter 2

Background

2.1 Consciousness

2.1.1 Historical background

Consciousness was long considered a taboo word in cognitive science, even when behaviourism, which substantially forbade any reference to mental states, was in decline (Seth, 2018). This can be illustrated by quotes such as: “Maybe we should ban the word for a decade or two” (Miller, 1962, p. 40), or “Consciousness is a fascinating but elusive phenomenon. It is impossible to specify what it is, what it does, or why it evolved. Nothing worth reading has been written on it” (Sutherland, 1989, p. 95).

Identifying the neural basis of consciousness in neural systems composed of billions of neurons is a serious challenge made worse by the presence of many confounding factors. For example, the fact that interrupting blood flow to the brain causes unconsciousness does not mean that blood flow should be considered a basis of consciousness. The goal is to isolate the necessary and sufficient conditions enabling conscious experience. Hence, the quest has been to find the minimal neuronal mechanisms jointly sufficient for any one conscious percept, so-called neural correlates of consciousness (NCC). It was pioneered by Francis Crick and Christof Koch whose seminal paper entitled “Towards a neurobiological theory of consciousness” brought renewed interest in consciousness research (Crick & Koch, 1990).

A few years later, Bernard Baars made substantial progress at explaining conscious access, the phenomenon whereby information is selected and rendered globally accessible in the brain for further processing and report. In “A Cognitive Theory of Consciousness” (Baars, 1993, p. 15), he gave the following operational definition of (access) consciousness based on the ability of subjects to report the content of their experience through their behaviour:

For now, we will consider people to be conscious of an event if (1) they can

say immediately afterwards that they were conscious of it and (2) we can independently verify the accuracy of their report. If people tell us that they experience a banana when we present them with a banana but not with an apple, we are satisfied to suppose that they are indeed conscious of the banana.

The notion of access consciousness has later raised the question of “why should global accessibility give rise to conscious experience?” constituting the hard problem in philosophy of mind (Chalmers, 1995, p. 8). Phenomenal consciousness, by opposition, denotes the experiential aspect of consciousness but remains largely unaddressed by neuroscience due to the lack of suitable methods to investigate it. However, the modelling methods I use here have been proven useful for giving a tentative neural account of phenomenological properties (Thagard & Stewart, 2014). As phenomenal consciousness is outside the scope of this work, the further use of the term consciousness will refer to the notion of access consciousness.

2.1.2 Minimal contrast paradigms

In order to find NCCs, paradigms allowing experimenters to manipulate consciousness with minimally different stimuli are of great help. Since the induced changes are minimal, neural activity in conditions of conscious perception can be compared with activity in unconscious conditions. Many such minimal contrast paradigms have been developed (Kim & Blake, 2005; Baars, 1993). In backward masking, a visual stimulus is not consciously seen when the time between its presentation and the presentation of a second distracting stimulus, called a mask, is too short (figure 2.1).

Together with the operational definition proposed by Bernard Baars which relies on the report of participants, minimal contrast paradigms provide a way of studying consciousness scientifically. This approach gave rise to one of the most important cognitive and neural theories of consciousness, as described in the next part.

2.1.3 Global workspace theory

Global workspace theory was introduced by Baars (1993) as an attempt at giving an account of the organisation of human cognition and describing the necessary and sufficient conditions for an information to be consciously experienced. It builds upon earlier ideas regarding the modular architecture of the mind (Fodor, 1983), while introducing a distinction between unconscious processors and a conscious global workspace which breaks the modularity of unconscious processing (figure 2.2).

Processors are systems composed of multiple processes involved in the fast and habitual transformation of mental representations, typically acquired through experience. An important characteristic of such processing is that it is informationally encapsulated,

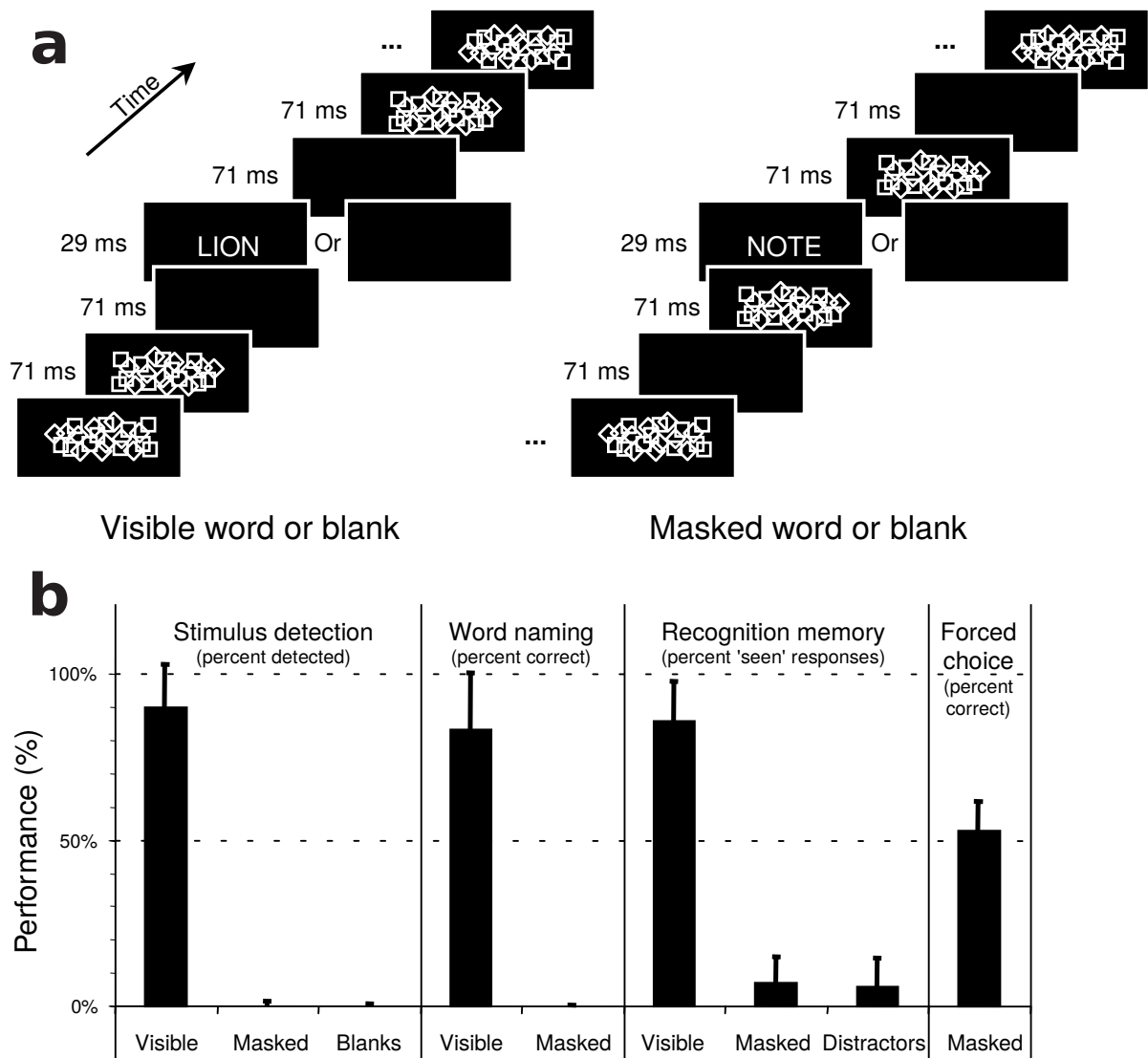


Figure 2.1: Backward masking experiment. When a stimulus is shortly followed by a mask (a; right), participants largely fail to detect it and report its identity (b). Taken from Dehaene et al. (2001).

or in other words, the information contained in processes is only accessible locally. A set of processes is defined as a processor when they all work together to perform a domain-specific function, such as understanding the meaning of a sentence or driving a car. Thus, they underlie semantic and procedural memory.

While information contained in processors is unconscious because of its encapsulation (Shanahan & Baars, 2005), the global workspace renders information conscious by making it accessible to all processors, including those enabling the report of experience through behaviour. This representational space thus enables system-wide communication. However, its capacity is limited and only some information can be granted access at a time, which explains the serial nature of conscious experience. In order to enter

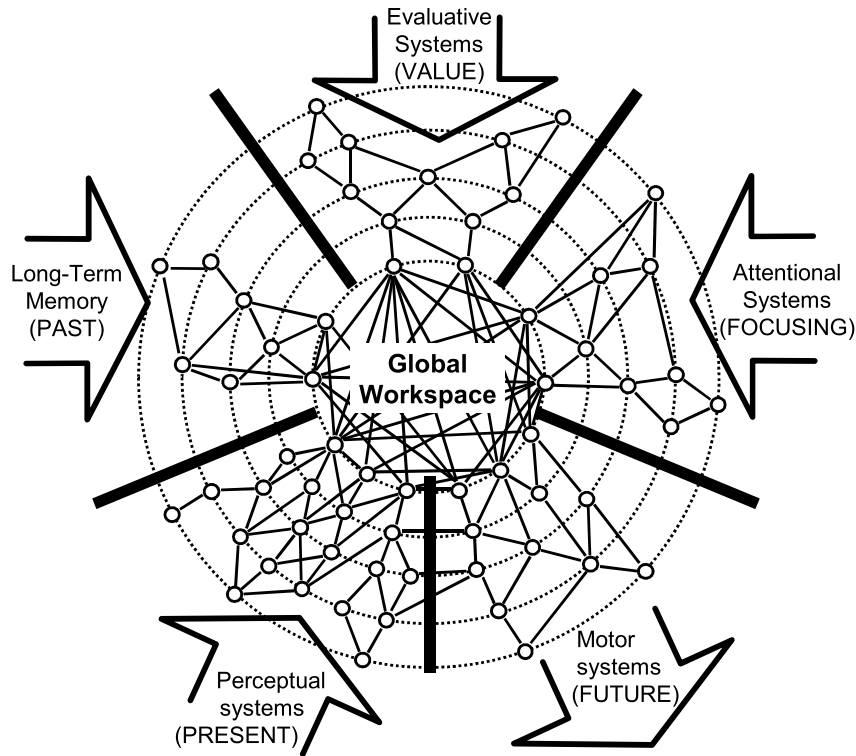


Figure 2.2: Global workspace architecture. Encapsulated systems composed of multiple processors compete for access to the global workspace. Conscious information can in turn be broadcast to all processors. Taken from Dehaene et al. (1998).

the workspace and become globally available, some piece of information must be selected through a competitive process. Dehaene et al. (2006) proposed a taxonomy to describe the different conditions under which a stimulus might be consciously perceived or not (figure 2.3). When the stimulus is too weak, its processing is confined to low levels of the hierarchy and does not compete for conscious access, regardless of the focus of attention. It is said to be subliminal. Alternatively, when it is sufficiently strong to access higher levels but fails to enter consciousness because it is not attended, the stimulus is said to be preconscious. However, it may be selected once the focus of attention shifts if its activation has been sustained. Only when it is both strong enough to enter the competition and the focus of attention is the stimulus granted access to the workspace. The subject then consciously perceives the stimulus and is able to use it for further processing or report it.

In several formulations of the theory, processors can collaborate and form a “coalition” to bring information in the conscious workspace (e.g. Baars et al., 2013; Shanahan, 2012; Baars & Franklin, 2009; Baars, 1993). This notion is shared by other accounts of consciousness (e.g. Crick & Koch, 2003). Because information is encapsulated, each processor can represent some feature of a percept without knowing that some information related to that same percept is being represented in other processors. The concept of

coalition implies that the global workspace is capable of treating these separate pieces of information as belonging to the same percept rather than only being able to select one of them. Understanding how the brain can achieve this requires a solution to the “binding problem” of how multifarious representations are integrated into single unified percepts. Since the workspace receives information from systems operating on different time windows (past, present and future ; figure 2.2), integration is both synchronic (at a moment in time) and diachronic (over time; Mashour et al., 2020).

In the next part, I describe the neural underpinnings of the global workspace theory. I also discuss its tentative neural models and why they fail at providing a satisfying account of serial processing and coalitions.

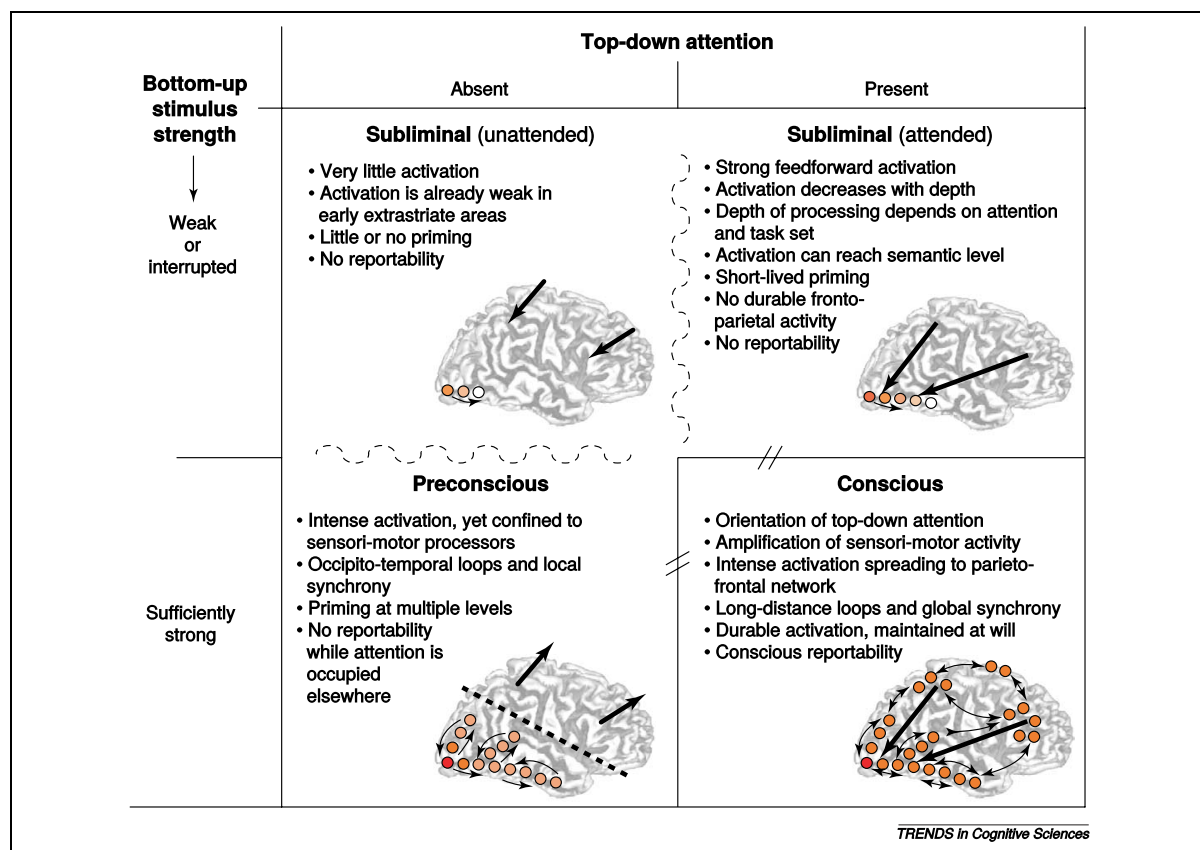


Figure 2.3: Taxonomy differentiating subliminal, preconscious and conscious processing. Shades of color indicate the amount of neural activation. Small arrows represent the interactions between areas, and large arrows the orientation of top-down attention. Dashed curves indicate a continuum of states while thick lines with separators indicate a sharp transition between unconscious and conscious states. Taken from Dehaene et al. (2006).

2.1.4 Neural implementations

During the last two decades, global workspace theory has been subject to a thorough investigation. In particular, the focus has been put on determining whether it accurately

describes the neural architecture underlying human consciousness. Many studies have found neural correlates consistent with a global workspace organization, leading to the formulation of a global neuronal workspace theory (Mashour et al., 2020; Dehaene, 2014; Dehaene et al., 2014; Dehaene & Changeux, 2011; Dehaene & Naccache, 2001). In the book “Consciousness and the Brain” (Dehaene, 2014), the following neural signatures of consciousness are described (figure 2.4):

1. The activity of many regions, including parietal and prefrontal circuits, is greatly amplified.
2. A late positive wave “ignites” as recorded on the scalp following the presentation of a seen stimulus by approximately 300 ms (P300 or P3 wave). This wave is not present when the stimulus remains unconscious¹.
3. The power of gamma oscillations increases starting at about 300 milliseconds after the presentation of a seen stimulus.
4. Distant brain areas synchronise.

As illustrated in figure 2.4, beta power has been shown to increase in conscious conditions (Gross et al., 2004). However, the opposite has been reported in other studies (Gaillard et al., 2009; Wyart & Tallon-Baudry, 2009). Results on gamma band power increase are much more convergent and therefore constitute a more reliable signature for conscious access (for a review see Dehaene & Changeux, 2011). These neural signatures lay the foundations of the global neuronal workspace theory. The workspace is hypothesised to be composed of a distributed set of neurons mainly located in the prefrontal and parietal cortices, and interconnected with long-range excitatory axons. These axons originate from pyramidal cells of layers 2 and 3 and project on either excitatory or inhibitory neurons (Dehaene et al., 1998). In comparison to sensory cortices whose connectivity is more local, layers 2 and 3 are thicker in dorsolateral prefrontal and inferior parietal cortices (figure 2.5). These numerous interconnections ensure that only a subset of neurons corresponding to a single conscious representation is active at a time while the other neurons are inhibited.

Access to this computational space is indexed by the P3 wave and gamma oscillations. As written in Dehaene (2014, p. 180), the positive wave can be explained by properties of workspace cells:

Now, the geometrical layout of the cells is such that, in the active ones, synaptic currents travel from the superficial dendrites toward the cells’ bodies. Because all these neurons are parallel to one another, their electrical

¹A first P3a wave, which can be evoked under unconscious conditions, can be distinguished from a later P3b component. Up to now, the latter has not been observed outside conscious conditions and is therefore thought to index conscious states.

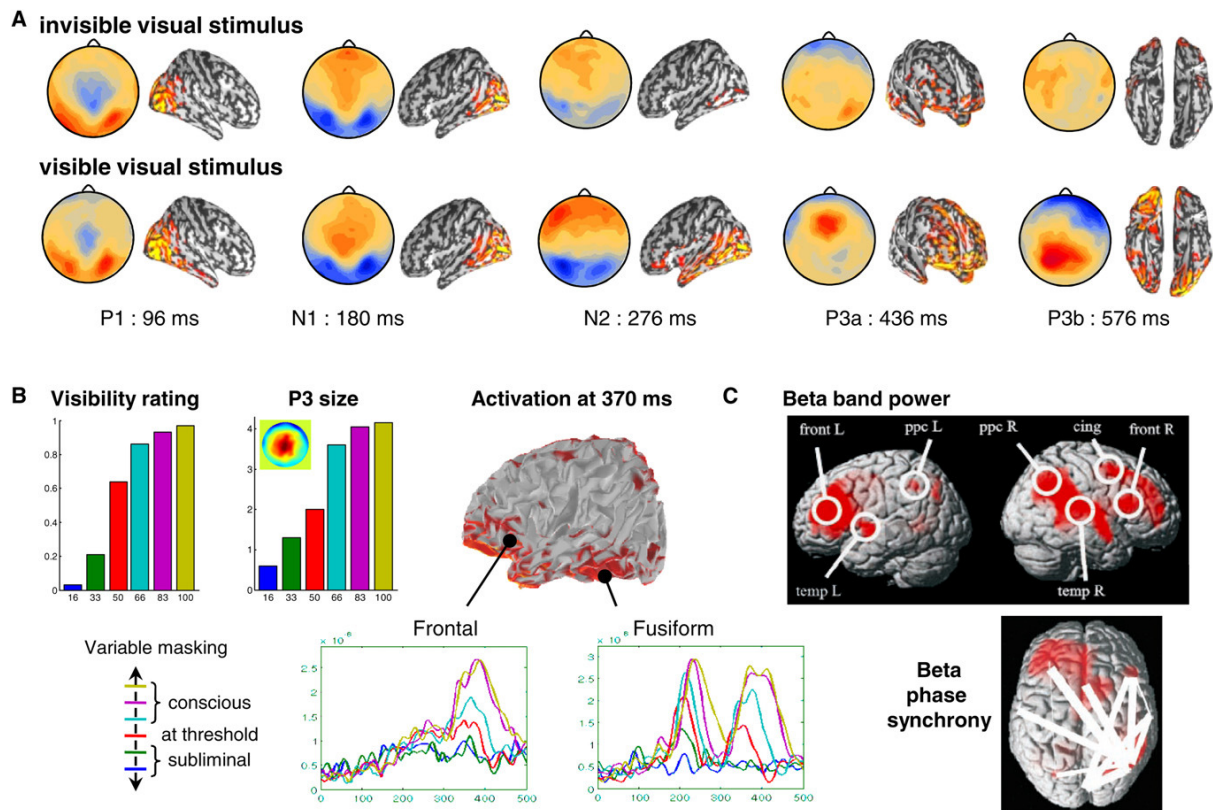


Figure 2.4: Signatures of conscious access. (A) Time course of event-related potentials following the presentation of a visual stimulus during an attentional blink experiment (another minimal contrast paradigm; Sergent et al., 2005). Early events in seen and unseen conditions do not differ. However, a N2 event (negativity around 200 ms) is amplified and the P3 event is present only when the stimulus is consciously perceived. (B) Visibility and brain activity as a function of the asynchrony between a visual stimulus and a mask in a backward masking experiment (Del Cul et al., 2007). Responses of high-level areas and visual experience follow a nonlinear function with respect to asynchrony, defining a threshold value for conscious access. Fusiform activation is first linear revealing unconscious processing, then nonlinear as information is amplified by top-down feedback in seen trials. (C) In an attentional blink study using magneto-encephalography, Gross et al. (2004) showed that induced power and phase synchrony increase in low beta frequencies on perceived trials. Taken from Dehaene & Changeux (2011).

currents add up, and, on the surface of the head, they create a slow negative wave over the regions that encode the conscious stimulus. The inhibited neurons, however, dominate the picture and their activity adds up to form a positive electrical potential. Because many more neurons are inhibited than are activated, all these positive voltages end up forming a large wave on the head, the P3 wave that we easily detect whenever conscious access occurs.

As the winning coalition is sent from a processor and broadcast to the others, the activity of multiple areas synchronises.

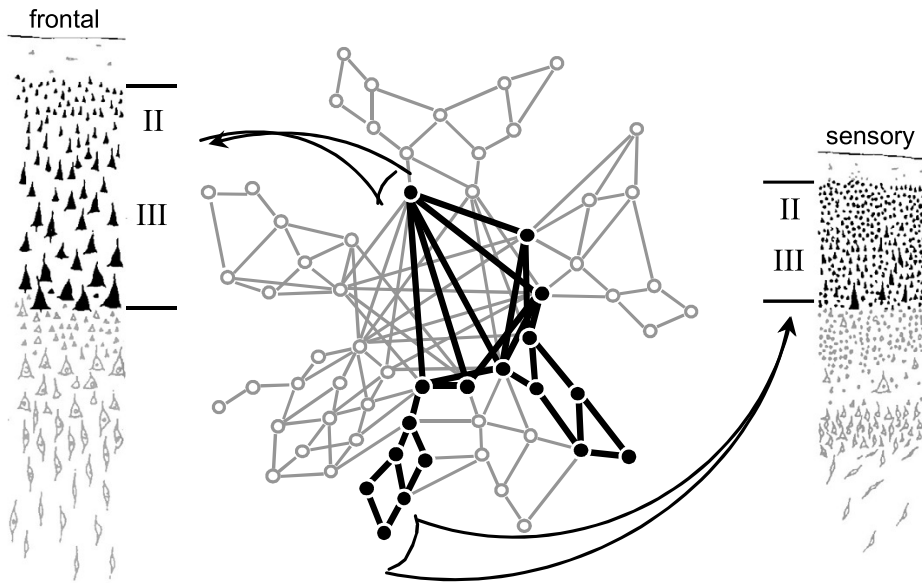


Figure 2.5: Thickness of layers 2 and 3 (containing large pyramidal neurons whose long axons project to distant regions) in frontal-type (left) and sensory-type cortices (right). Taken from Dehaene et al. (1998).

2.1.5 Global neuronal workspace simulations

Many computational models of the theory have been proposed. Since an exhaustive review is outside my scope of work, I briefly describe those based on spiking neurons. Dehaene et al. (2003) proposed a neural model in line with anatomical and physiological data with cortical columns of 3 layers connected to a thalamic network. This model exhibits P3-like ignition, gamma oscillations and distant synchronisation. Moreover, the all-or-none dynamics exhibited by the model account for subjective reports in the attentional blink paradigm. An extension of the model was later used to simulate stochastic spontaneous activity of the workspace and how it may explain inattention blindness (Dehaene & Changeux, 2005). Another research team proposed a model accounting for both global workspace competition and broadcast (Shanahan, 2008). Interestingly, it was shown to exhibit the same behaviour with stochastic wiring (Connor & Shanahan, 2010). Finally, a model developed by Zylberberg et al. (2010) reproduces the psychological refractory period and attentional blink surprisingly well. Although the paper makes no explicit mention of global workspace theory, the model is composed of sensory and motor modalities competing for access to a central routing bottleneck which exhibits all-or-none dynamics. However, it cannot account for the richness of human sequential processing. In fact, it can only perform two competing tasks by executing a single routing command (information is sent from one region to another). In order to perform multi-step tasks in which the result of each step is used as input for the next step, routing mechanisms must be much more flexible. Furthermore, the model does not scale well, as each symbol

of each sensory modality projects to one selective population in the router. Hence, it cannot explain how multiple processors are integrated and greatly differs from the global neuronal workspace in that aspect.

To my knowledge, and as pointed out by Shanahan (2012), none of the proposed neural models addresses the coalition formation process. This is due to the fact that no elaborate link is drawn between neural activity and cognitive representations in these experiments. In that context, the binding problem cannot be addressed. Moreover, although global workspace theory is thought to support flexible serial processing (Mashour et al., 2020; Zylberberg et al., 2011), the models cannot perform cognitive tasks composed of multiple steps (but see Fountas et al., 2011).

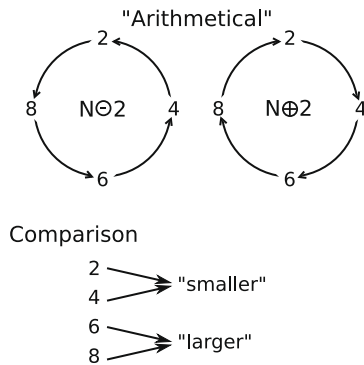
2.2 Flexible sequential processing

The “brain is a computer” analogy is not popular in modern cognitive science, as neurons work in parallel and are not synchronised on a central clock like computers. Yet somehow these disparate cells give rise to serial behaviour, for example “in the process of computing a real number” (Turing, 1936, p. 231). How the brain does this is still not well understood. Of course, individual processors can learn to perform a sequence of operations. However, the ability to flexibly route information and thus break the modularity of processors organisation is required to perform new multi-step tasks. Although computational views of the mind are considered outdated, Alan Turing’s universal computing machine, the ancestor of modern serial computers, was inspired from human cognition and may in fact function similarly to conscious processing Zylberberg et al. (2011). This is best demonstrated by a study from Sackur & Dehaene (2009), in which participants were shown a digit and asked to perform different mathematical tasks (figure 2.6). The digit was a natural integer $N \in \{2, 4, 6, 8\}$. In the simple task, participants were asked to report whether the stimulus N was smaller or larger than the fixed reference 5. In two other tasks, participants were asked to first perform an arithmetic operation on N before comparing the result of that operation to 5. In the chained-addition task, the first operation to perform was a modified addition $N \oplus 2$, while in the chained-subtraction task, the operation was a modified subtraction $N \ominus 2$. These operations differ from their ordinary counterpart $+$ and $-$ in that a cycling rule is introduced, as figure 2.6 illustrates. This rule brings the advantage that the result of the operations remain within the set of input stimuli $\{2, 4, 6, 8\}$. Operations \oplus and \ominus can thus be defined as follows:

$$N \oplus 2 = \begin{cases} N + 2, & N \in \{2, 4, 6\} \\ 2, & N = 8 \end{cases} \quad (2.1)$$

$$N \ominus 2 = \begin{cases} N - 2, & N \in \{4, 6, 8\} \\ 8, & N = 2. \end{cases} \quad (2.2)$$

Operations



Tasks

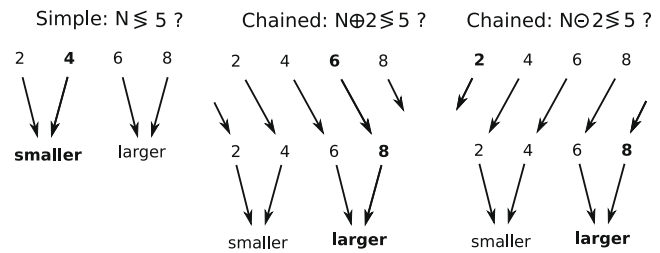


Figure 2.6: Operations and tasks to perform. The cycling rule shown on the left ensures that digits remain within the set of input stimuli. Taken from Sackur & Dehaene (2009).

2.2.1 Experiment 1: chronometric exploration

The first experiment aimed at determining whether participants would take more time to perform the composite tasks ($N \oplus 2 \leq 5?$ and $N \ominus 2 \leq 5?$) in comparison to the simple task ($N \leq 5?$), and if the operation chaining was performed in an optimal serial way. In order to perform one of the composite tasks, an ideal serial computing machine would first perform the arithmetic operation, wait for that operation to be done, then and only then perform the comparison operation using the result of the first operation. Model 1 in figure 2.7 is a depiction of such a machine.

The results of this experiment are shown in figure 2.8. A distance effect was observed in simple trials such that the greater the distance between the input stimulus and the reference five was, the faster the comparison operation proceeded and the less errors were made. This effect is well known in psychology (e.g. Moyer & Bayer, 1976). Results also indicate that participants took 85 ms longer to answer and made 4.3% more mistakes in chained trials than in simple trials. In addition, there was a clear difference between congruent trials in which both the input stimulus and the result of the arithmetic operation fell on the same side of five (e.g. $N = 2, N \oplus 2 \leq 5?$), and incongruent trials in which

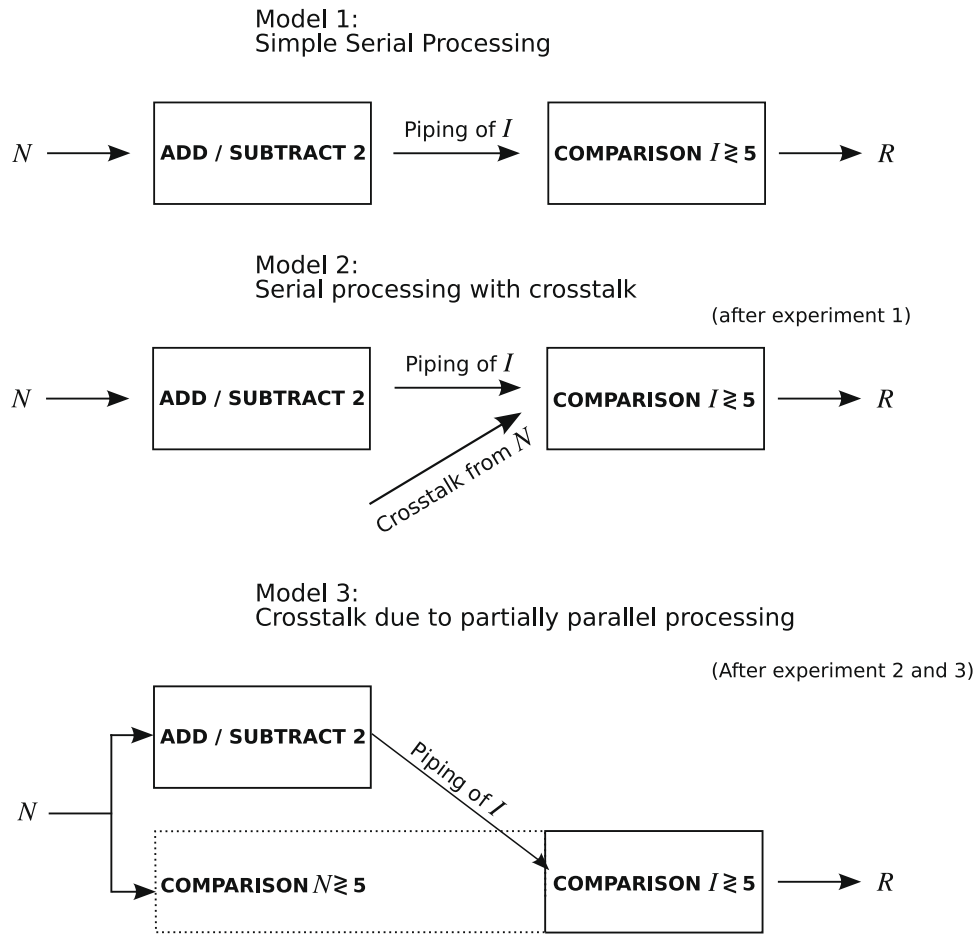


Figure 2.7: Candidate models for chaining operations. Taken from Sackur & Dehaene (2009).

this is not the case (e.g. $N = 4, N \oplus 2 \leq 5$?). Response times and error rates were higher in incongruent conditions.

The simple serial model predicts that there should be a distance effect in chained trials with respect to the image of the arithmetic operation. However, no such effect was found in the statistical analyses. There was no significant interaction between response times or error rate and the distance between five and the input stimulus either. This combined with the observation of an unpredicted congruence effect are evidence against the simple serial model. In fact, these results suggest that information about the input stimulus is used in the comparison operation, which leads to more errors and slower response times in incongruent trials and potentially compensates the image distance effect.

Sackur & Dehaene (2009) proposed two alternative models with crosstalk from the input to the comparison (figure 2.7). Like the simple serial model, Model 2 executes the operations sequentially but takes both the image of the arithmetic operation and the input stimulus as inputs for the comparison operation. The other alternative model is only partially serial. Both operations are performed in parallel when the input digit is presented, and the result of the arithmetic operation is piped to the comparison operation.

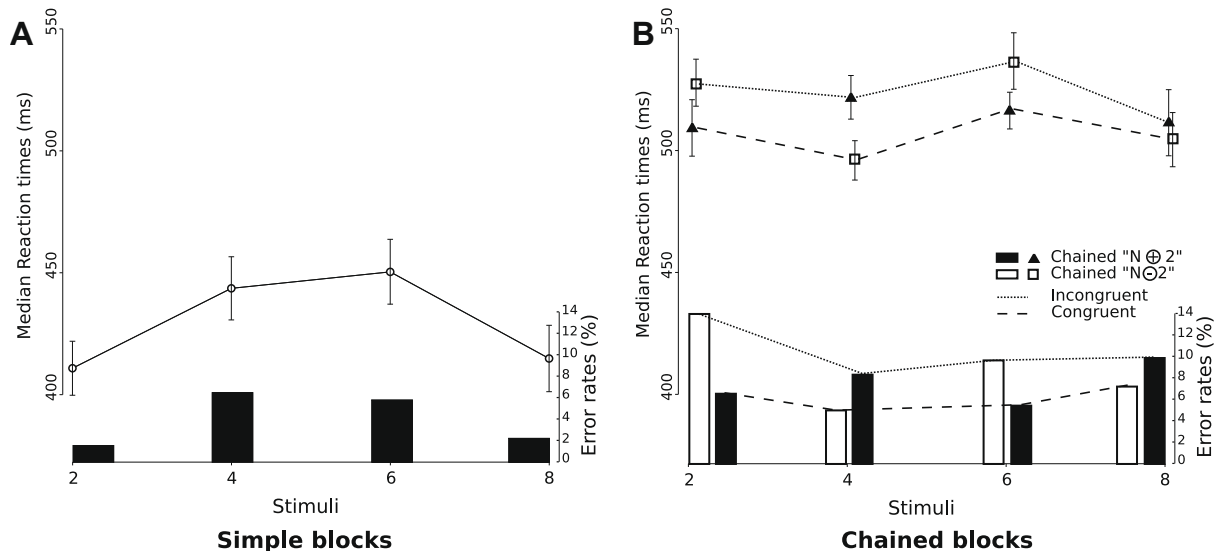


Figure 2.8: Median response times and error rates in simple and composite tasks. When the result of the first operation does not fall on the same side of five than N , the trial is said to be incongruent (dotted lines). When it does fall on the same side, the trial is congruent (dashed lines). Error bars represent plus and minus one standard error of the mean after removal of the subject means. Taken from Sackur & Dehaene (2009).

As both models can account for the results of this experiment, Sackur & Dehaene (2009) performed other experiments in order to investigate the time course of the operations and differentiate between the two.

2.2.2 Experiment 3: priming

The results of the second experiment favour Model 3 but are less clear than those of Experiment 3 so I will not discuss them here. The design of the priming experiment differs slightly from the procedure described above (figure 2.9). A first digit was presented as a prime, followed by a target digit to be compared to five with a variable stimulus onset asynchrony (SOA). Participants were asked to use the prime to answer faster, as in 66% of trials, the target would be $T = P \oplus 2$, $T = P \ominus 2$, or $T = P$ in the chained-addition, chained-subtraction and simple tasks respectively. The proportion of informative primes (66%) was not given to the subjects.

If Model 2 underlies the execution of this experiment, congruence effects independent from SOA should be observed as the input to the comparator processor should be a mixture of the computed image, the target and the prime. Model 3 makes different predictions regarding the results of this experiment. A first automatic prime-target congruence effect should be observed at short SOAs, reflecting the start of the comparison operation with the prime. A later strategic image-target congruence effect should also be observed as a result of the execution of the arithmetic operation and the piping of its

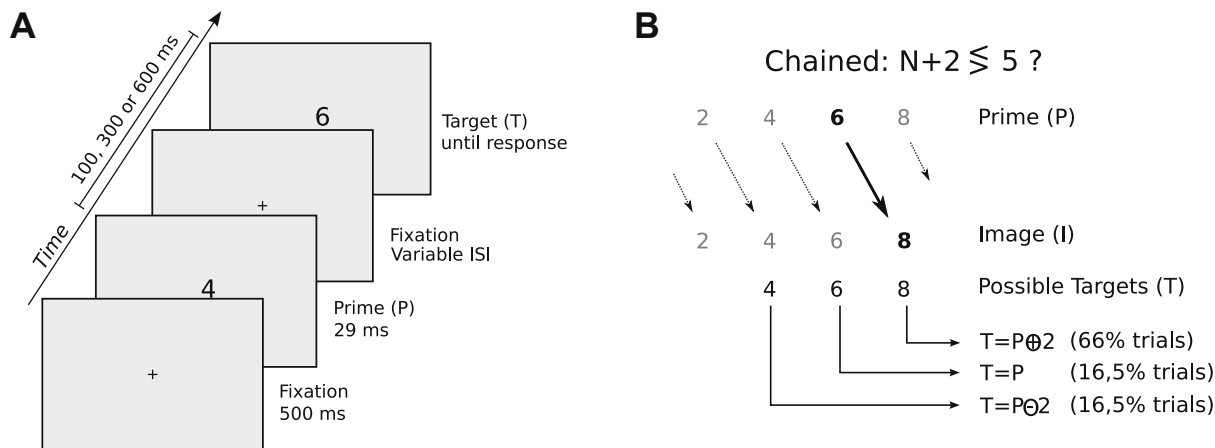


Figure 2.9: Priming experiment. (A) Experimental procedure. A prime digit P precedes a target T by a variable delay. (B) Description of chained-addition trials. In 66% of trials, the target is the result of the first operation to perform (i.e. image I). Taken from Sackur & Dehaene (2009).

result to the comparator. These two priming effects dependent on SOAs were observed, providing strong evidence against Model 2 (figure 2.10). Hence, this study suggests that humans perform serial tasks with interference from partially parallel processing. Since piping information once the arithmetic operation is executed requires cognitive control, the relationship of this cognitive architecture with consciousness was investigated in a last experiment.

2.2.3 Experiment 4: backward masking

The last experiment uses backward masking, the minimal contrast paradigm described in section 2.1.2. Two additional tasks were introduced (figure 2.11). The naming task was to identify and report the input stimulus. The arithmetic task was to report either $N \oplus 2$ or $N \ominus 2$. The other tasks were the same as those in the previous experiments. The performance of participants in all task types is shown in figure 2.11 as a function of stimulus visibility. Even when the stimulus was not consciously perceived, participants performed significantly better than chance in the naming, arithmetic and simple tasks. However, their performance was at chance for composite tasks.

This study shows that the serial processing required to perform new multi-step tasks is not a natural mode of operation for the human brain, as confirmed by more recent reports (Fan et al., 2012b,a). In fact, steps that should be performed at later stages start on the input stimulus in parallel with the first operations to perform. Effortful and high-level cognitive control enabled by conscious perception is required to take charge of sequential processing and information flow. Furthermore, sustained activity is observed in left temporal, left parietal and left occipital areas when performing chained tasks

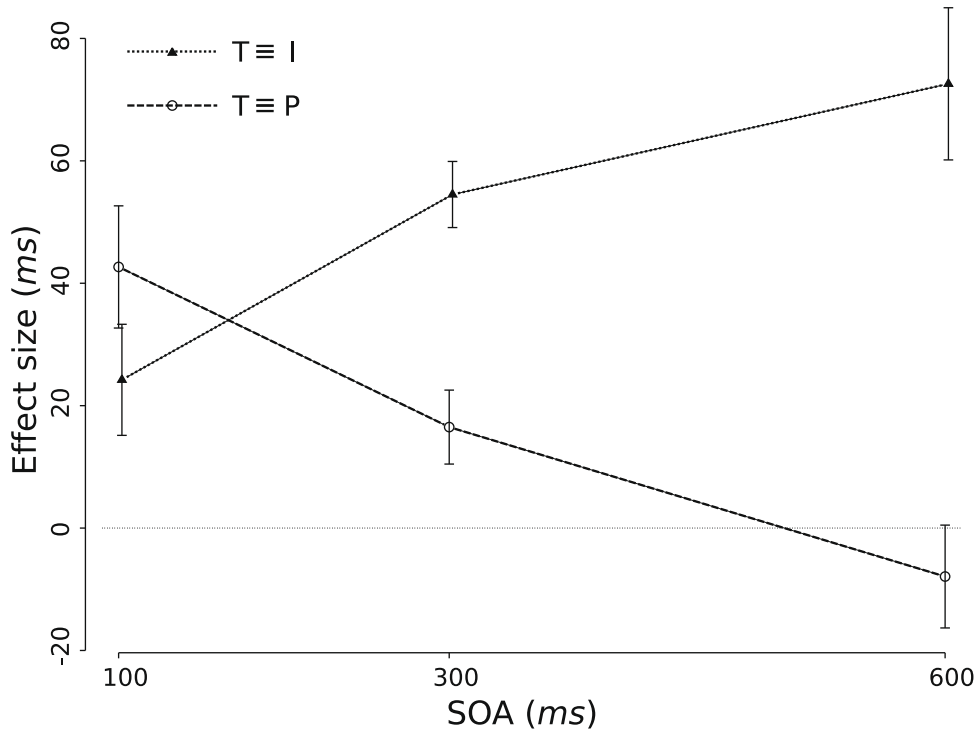


Figure 2.10: Congruence priming effects within chained blocks as function of stimulus onset asynchrony. Prime-target congruence effects are calculated as RTs for non prime congruent trials (the prime and the target do not fall on the same side of 5) minus RTs for prime congruent trials ($T \equiv P$: they are on the same side). Similarly, image-target congruence effects are calculated as RTs for non image congruent trials (the image and the target do not fall on the same side of 5) minus RTs for image congruent trials ($T \equiv I$: they are on the same side). Error bars represent plus and minus one standard error of the mean after removal of the participant mean. Taken from Sackur & Dehaene (2009).

(Fan et al., 2012a). As Sackur & Dehaene (2009) pointed out, the functions that global workspace theory attributes to consciousness are those required for the mode of operation enabled by conscious perception in this study. With a global workspace architecture, multi-step tasks can be performed as follows: (1) the stimulus accesses the workspace, (2) it is then sent to the processor responsible for performing the first operation, (3) the result accesses the workspace, (4) it is then sent to another processor to perform the next step, etc. Congruence effects in the first experiments and slightly higher than chance performance in single-step tasks of the masking experiment can be explained by partial communication between at least some processors involved in the tasks. Hence, the global workspace architecture can theoretically account for the results of Sackur & Dehaene (2009).

The aim of this thesis is to build a spiking neural model compatible with global workspace theory and Model 3 to assess whether it is able to (1) perform multi-step tasks and (2) reproduce the results of Experiment 1. Since experiments 2,3 and 4 were meant to investigate the cognitive architecture underlying the execution of the first experiment,

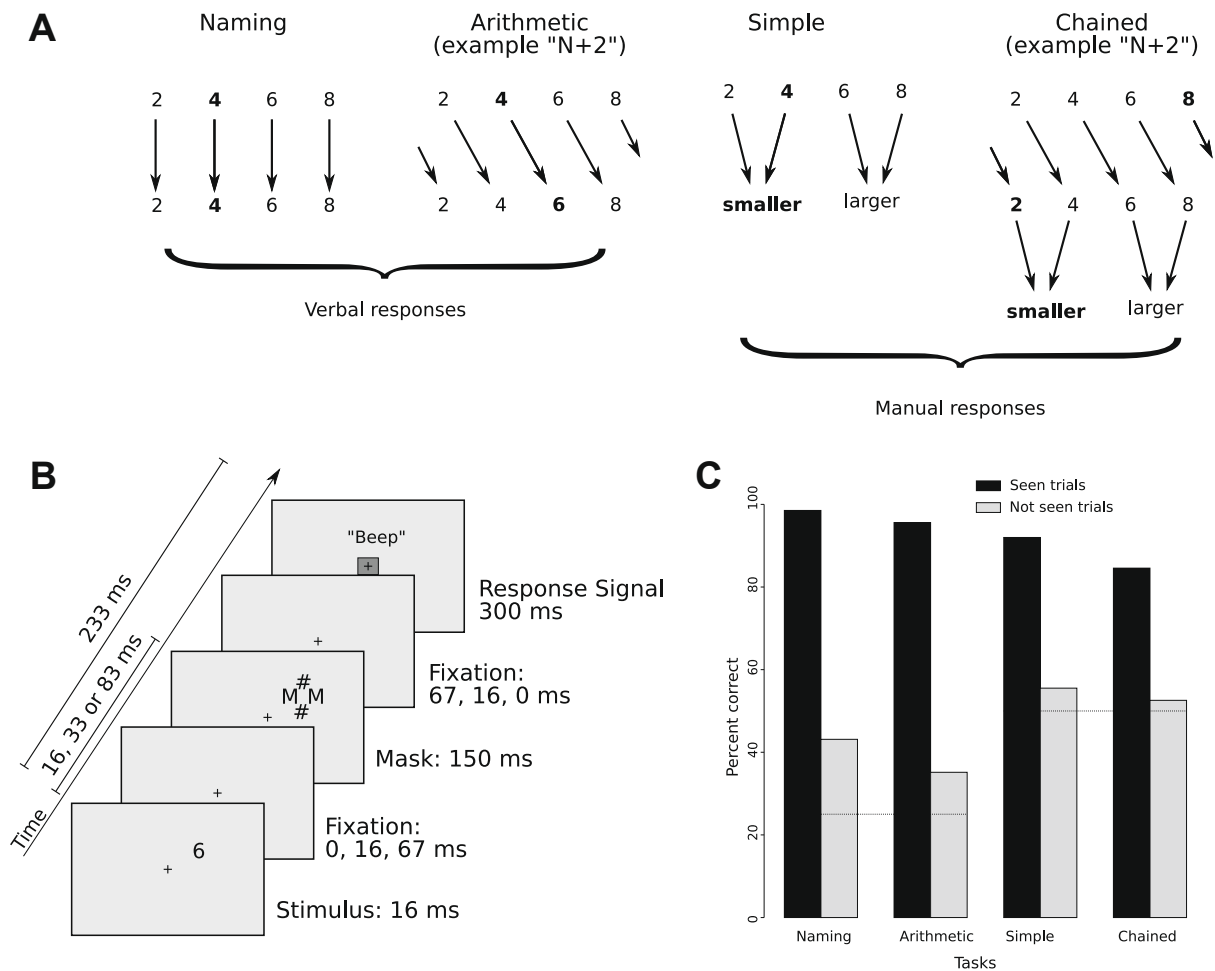


Figure 2.11: Masking experiment. (A) Tasks to perform. Answers to naming and arithmetic tasks are reported verbally, and answers to simple and composite tasks are reported by pressing one of two buttons like in the previous experiments. (B) Experimental procedure. The mask is inserted at a variable time between the stimulus and a sound calling for the response. (C) Performance of participants as a function of stimulus visibility. Dotted lines indicate chance levels in all task types. Taken from Sackur & Dehaene (2009).

reproducing the first experiment with a model that is by construction compatible with the results of the others is considered satisfactory. Of course, reproducing the other experiments is important too in order to fully test the model, but this is left for future research.

Chapter 3

Methods

The frameworks implemented in Nengo (Bekolay et al., 2014), the software I use for creating and simulating the neural models, provide a way of creating spiking neural models capable of performing high-level cognitive tasks. In this chapter, I describe these frameworks and a way of optimising their parameters.

3.1 Neural Engineering Framework

The three principles of the Neural Engineering Framework (NEF) described below were introduced by Eliasmith & Anderson (2003). They provide a way to implement cognitive models specified in the format of dynamical systems and vector operations in networks of spiking neurons. As described in section 3.2, the NEF is also used with a certain type of vectors, called semantic pointers, in order to represent and manipulate semantic information.

3.1.1 Representation

The first principle of the NEF postulates that a population of neurons, called an “ensemble”, represents a real-numbered vector \mathbf{x} of \mathcal{D} dimensions. Each neuron i in the population has a gain parameter α_i , a bias current b_i and a nonlinearity G defined by the neural model. In this work, all neurons are leaky integrate-and-fire (LIF), as this neural model provides a good trade-off between biological plausibility, interpretability and computational cost. Thus, G converts the internal current of the neuron into a neural spike train. The crucial element to map between the vector \mathbf{x} being represented and the activity of neuron i is the encoding vector \mathbf{e}_i , which corresponds to the direction in the vector space that makes the neuron fire the fastest. The spike train a_i is computed as follows:

$$a_i(t) = G(\alpha_i \mathbf{e}_i \cdot \mathbf{x}(t) + b_i) \tag{3.1}$$

Equation 3.2 describes how an approximation of \mathbf{x} can be decoded back from the spike trains. Neural activity is first convolved with an exponentially decaying synaptic filter $h_\tau(t) = \exp(-t/\tau)\tau^{-1}$ with time constant τ , then multiplied by a decoding vector \mathbf{d}_i .

$$\hat{\mathbf{x}}(t) = \sum_i \mathbf{d}_i (\mathbf{a}_i * h)(t) \quad (3.2)$$

In order to find the linear decoders \mathbf{d}_i , \mathbf{x} values (called evaluation points) are randomly sampled and the reconstruction error E is minimised through regularised least squares optimisation.

$$E = \sum_x \|\mathbf{x} - \hat{\mathbf{x}}\|^2 \quad (3.3)$$

3.1.2 Computation

The second principle describes how neurons are connected in order to implement the desired transformations. The weight matrix is defined as the outer product of the encoders and the decoders $w_{ij} = \mathbf{e}_i \times \mathbf{d}_j$. Equation 3.1 can then be rewritten with a weighted summation of postsynaptic potentials.

$$\mathbf{a}_i(t) = G \left[\sum_j \alpha_j w_{ij} (\mathbf{a}_j * h)(t) + b_i \right] \quad (3.4)$$

By replacing \mathbf{x} and $\hat{\mathbf{x}}$ with $f(\mathbf{x})$ and $\hat{f}(\mathbf{x})$ in equation 3.3, decoders for a connection can be solved to implement arbitrary transformations:

$$E = \sum_x \|f(\mathbf{x}) - \hat{f}(\mathbf{x})\|^2 \quad (3.5)$$

3.1.3 Dynamics

In order to account for the variety of neural systems, a theory of neural coding must also explain the role of feedback connections. Using the Computation principle, neural populations can be connected to themselves. When such recurrent connections are introduced, state equations of the following form can be implemented:

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) \quad (3.6)$$

$$\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \quad (3.7)$$

where $\mathbf{x}(t)$ is the state vector, $\mathbf{u}(t)$ the input, and $\mathbf{y}(t)$ the output. The system behaviour is determined by the dynamics matrix \mathbf{A} , input matrix \mathbf{B} , output matrix \mathbf{C} , and the

feedthrough matrix \mathbf{D} . Recall that synaptic filter $h_\tau(t)$ of time constant τ drives the neuron dynamics. Thus, the state equations are implemented by

$$\mathbf{x}(t) = h_\tau(t) * [\mathbf{A}'\mathbf{x}(t) + \mathbf{B}'\mathbf{u}(t)] \quad (3.8)$$

where $\mathbf{A}' = \tau\mathbf{A} + \mathbf{I}$ and $\mathbf{B}' = \tau\mathbf{B}$ are found with the Laplace transform (Eliasmith & Anderson, 2003). Nonlinear dynamics can also be implemented using the Computation principle.

3.2 Semantic Pointer Architecture

The Semantic Pointer Architecture (SPA) is a cognitive architecture that makes use of the notion of vector-symbolic architecture, a family of approaches that has been proven useful for modelling high-level cognition (Eliasmith, 2013; Eliasmith et al., 2012; Gayler, 2004). More precisely, SPA partly implements Holographic Reduced Representations (Plate, 1995) in spiking neurons using the NEF. The approach is the following: a symbol (or concept), called a semantic pointer and denoted by a bold capital letter, is defined as a unit vector in a high-dimensional space. Four operations are used to manipulate and combine semantic pointers to form new ones:

1. *Scaling*: the scalar multiplication $\mathbf{A} = c\mathbf{B}$ gives a vector with the same direction as \mathbf{B} , with a magnitude scaled by a factor of c .
2. *Comparing*: the dot product $\mathbf{A} \cdot \mathbf{B}$ computes how close the vectors of semantic pointers \mathbf{A} and \mathbf{B} lie in the vector space. This provides a measure of semantic similarity.
3. *Superposition*: the addition of pointer vectors $\mathbf{A} = \mathbf{B} + \mathbf{C}$ gives a vector corresponding to a semantic pointer \mathbf{A} , similar to both \mathbf{B} and \mathbf{C} .
4. *Binding*: the binding of two pointers $\mathbf{A} = \mathbf{B} \circledast \mathbf{C}$ gives a pointer \mathbf{A} dissimilar to both \mathbf{B} and \mathbf{C} . \circledast denotes the circular convolution. To retrieve a pointer from the binding result \mathbf{A} , the circular convolution can be used with the approximate convolutive inverse of one of the original pointers. For example, $\mathbf{B} \approx \mathbf{A} \circledast \mathbf{C}^{-1}$ with \mathbf{C}^{-1} the approximate inverse of \mathbf{C} .

Note that all operations preserve the dimensionality of the vectors (\mathcal{D}). SPA also makes use of cortico-basal ganglia-thalamocortical loops for cognitive control. Basal ganglia are thought to implement winner-take-all computation, or in other words select the action with highest input utility (Redgrave et al., 1999). The model proposed by Gurney et al. (2001) was implemented in spiking neurons using the NEF (figure 3.1;

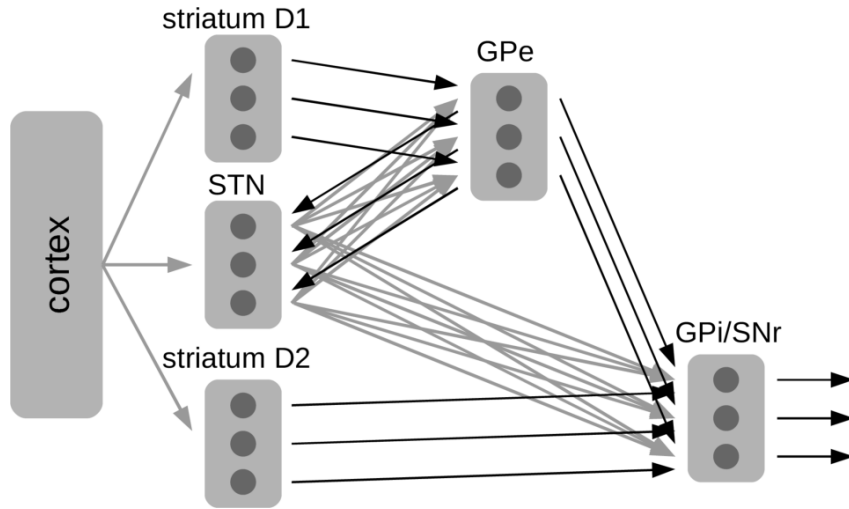


Figure 3.1: Basal ganglia model based on Gurney et al. (2001) with three possible actions. Light lines are excitatory connections. Dark lines are inhibitory. Taken from Stewart et al. (2010a).

Stewart et al., 2010b). Combined with thalamic and cortical components, it works as a production system (Stewart et al., 2010a). Thus, it can recognise cortical patterns, and consequently send information to cortical populations or route information from one cortical population to another. Figure 3.2 shows the implementation of the following toy rules:

$$\begin{array}{ll}
 \text{if} & \mathbf{x}_1 \cdot \mathbf{A} \\
 \text{is maximal, then} & \mathbf{B} \gg \mathbf{x}_2, \\
 & \mathbf{x}_1 \gg \mathbf{x}_2. \quad (\text{i.e. } \mathbf{x}_1 \otimes \mathbf{I} \gg \mathbf{x}_2) \\
 \text{if} & \mathbf{x}_1 \cdot \mathbf{B} \\
 \text{is maximal, then} & \mathbf{A} \gg \mathbf{x}_1.
 \end{array}$$

3.3 Parameter optimisation

In order to fit the model described in the next chapter to human behaviour¹, an evaluation function is minimised with respect to some parameters. The choice of function is crucial and depends on the scope of the project. In Sackur & Dehaene (2009), the median correct response times and error rates are used. We can denote them by $\widetilde{\text{RT}}$ and ER , respectively. These metrics must be indexed by the experimental condition (and parameters in the case of the model). Furthermore, subscripts h and m are introduced to distinguish between

¹For unknown reasons, the figures and statistical analyses reproduced from Sackur & Dehaene (2009) differ from those in the original publication, despite consultation with the first author.

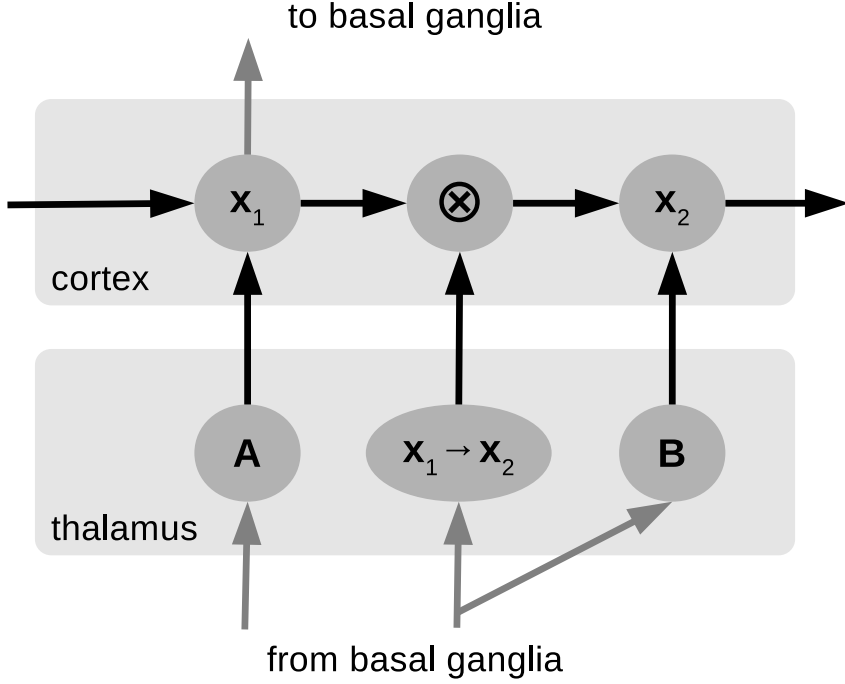


Figure 3.2: Thalamus and cortex model. Dark connections convey semantic pointers and light ones convey scalar values. \mathbf{x}_1 and \mathbf{x}_2 are cortical populations representing semantic pointers. \otimes are neurons computing the binding operation. Thalamic populations represent scalar values and project semantic pointers to cortical populations.

human and model behaviours. Thus, $\widetilde{\text{RT}}_m(t, N, \theta)$ is the median correct response time taken by the model created with parameters θ to answer trials of task t and input N . Parameter θ_s is the delay of unmodelled sensorimotor processing. Since θ_s can simply be added to the median response time, the model needs not be simulated multiple times when only this delay varies. The root-mean-square error is not an appropriate metric because it would penalise deviations from empirical means despite variation in human data. Instead, the fitness error function is defined as follows:

$$\mathcal{L}(\theta, \theta_s) = \sqrt{\frac{1}{12(\alpha_{\text{RT}} + \alpha_{\text{ER}})} \sum_{t \in \{\text{SIMPLE}, \text{CHAINED_ADD}, \text{CHAINED_SUB}\}} \sum_{N \in \{2, 4, 6, 8\}} \alpha_{\text{RT}} \Delta_{\text{RT}}^2 + \alpha_{\text{ER}} \Delta_{\text{ER}}^2} \quad (3.9)$$

$$\Delta_{\text{RT}} = \widetilde{\text{RT}}_h(t, N) - (\widetilde{\text{RT}}_m(t, N, \theta) + \theta_s) \quad (3.10)$$

$$\Delta_{\text{ER}} = \text{ER}_h(t, N) - \text{ER}_m(t, N, \theta) \quad (3.11)$$

For each parametrization θ , 10 participants² are simulated performing all experimental conditions twice (in random order). Since \mathcal{L} is computationally expensive to evaluate, Bayesian optimisation is a good approach for minimising it. The function is approxi-

²Different participants are simulated by setting the seed of the random generator used in various parts of the model (e.g. for generating atomic semantic pointers or sampling evaluation points).

mated using a Gaussian process with a Matérn kernel which assumes the function values follow a multivariate gaussian. This process is implemented using scikit-optimize (Head et al., 2018). Once the simulations have run with parameters θ , it is easy to compute \mathcal{L} with variable θ_s . Therefore, the function is minimised with respect to delay θ_s inside each function evaluation of the Bayesian optimisation algorithm. This is done using the bounded-limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (Byrd et al., 1995) implemented in SciPy (Virtanen et al., 2020).

Chapter 4

Model

In this chapter, I describe the model for chaining mental operations¹. Although the processors depicted in figure 4.1.A are designed specifically for performing the tasks of Sackur & Dehaene (2009), the general architecture follows global neuronal workspace theory and could be used to chain information between virtually any processors (figure 4.1.B). Three processors perform the mathematical operations: *Compare*, *Add* and *Subtract*. Router components control information flow via product networks that weight the input to and output from the global workspace.

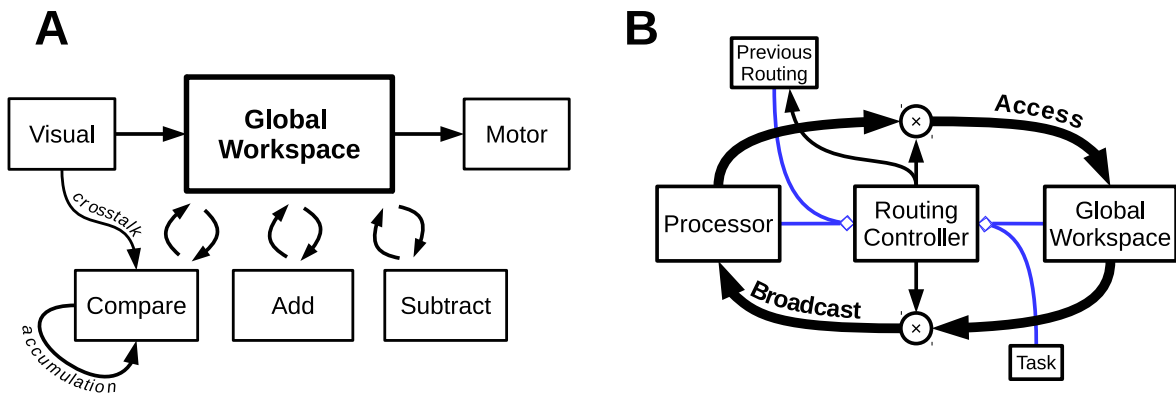


Figure 4.1: Model architecture. (A) Processors used for the tasks and global workspace. The crosstalk connection is included to reproduce the stimulus-image congruence effect. The comparison operation is performed by an accumulator-to-threshold to reproduce both the congruence and distance effects. (B) Interactions between a processor and the rest of the model. Blue connections are used to compute production rules utilities.

¹The code is available at <https://github.com/HugoChateauLaurent/conscious-chaining>.

4.1 Processors

As defined in section 2.1.3, processors are associated with long-term semantic and procedural memory. They can be composed of virtually any set of processes. Thus, a processor is any neural network that has learned (or was inherited) to perform domain-specific transformations and communicates with the global workspace. Three types of processors are used to model the task described by Sackur & Dehaene (2009). I describe them below.

4.1.1 Sensory and motor

A processor is referred to as perceptual when it receives direct sensory inputs, and motor when directly involved in the preparation or execution of movement. For simplicity, and because the precise modelling of visual processing and motor control is not the focus of the present work, *Visual* and *Motor* need not perform any computation. Here, *Visual* is given and outputs the semantic pointer corresponding to the task input. The behavioural response is directly decoded from what *Motor* receives from the workspace.

4.1.2 Hetero-associative memory

A second type of processors uses a hetero-associative memory which associates a value to a key. It can be implemented with a one-layer-deep network that transforms its input through synaptic connections (figure 4.2). In the layer, a neural ensemble p_j^i belonging

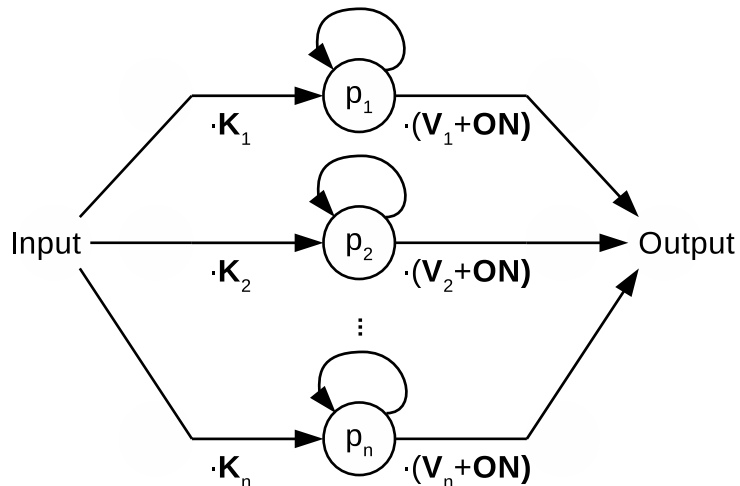


Figure 4.2: Hetero-associative memory. Ensembles $p_i, i = 1, 2, \dots, n$ represent the activation of their associated key semantic pointer \mathbf{K}_i . The input is translated to a scalar value by computing the dot product of the input and the semantic pointer to recognise. Similarly, the output is the product of activation p_i and pointer $\mathbf{V}_i + \mathbf{ON}$.

to the i -th processor is responsible for recognising a semantic pointer $\mathbf{K}_j \in \mathcal{K}$. \mathcal{K} is the

set of keys to be recognised by the processor. $\mathbf{p}^{i\leftarrow}$ is the input to processor i (in semantic pointer space). The dynamics are given by the following equation, prescribed by the Dynamics principle:

$$p_j^i(t) = \underbrace{h_{0.001}(t) * [0.001 \mathbf{p}^{i\leftarrow}(t) \cdot \mathbf{K}_j]}_{\text{input}} + \underbrace{h_{0.01}(t) * [(0.7 \times 0.01 + 1) p_j^i(t)]}_{\text{feedback}}. \quad (4.1)$$

With feedback weight 0.7, information decays in a way consistent with unconscious processing. In order to enforce the constraint $p_j \geq 0$, encoders are chosen in such a way that neurons only respond to positive inputs. From now on, the same applies to any network enforcing such a constraint. The scalar output of the ensemble is multiplied by semantic pointer $\mathbf{V}'_j = \mathbf{V}_j + \mathbf{ON}$, where $\mathbf{V}_j \in \mathcal{V}$ is a value pointer associated with key \mathbf{K}_j and \mathbf{ON} indicates the amount of preconscious information in the processor to the router described in section 4.4.

Processors *Add* and *Subtract* associate the semantic pointers of results $N \oplus 2$ and $N \ominus 2$ (superposed with \mathbf{ON}), respectively, to the semantic pointer of input N . This is probably a simplification of what happens in biological brains. In fact, it is unlikely that the brain uses one network per number to add or subtract. However, nothing in the behavioural study indicates that a more complex network is needed to reproduce human performance. Furthermore, Stewart et al. (2017) have demonstrated that a single layer network can compare any two numbers, and the same approach could be used for performing arithmetic operations. In fact, I use their approach for modelling a third type of processors.

4.1.3 Accumulation-to-threshold

Compare is an accumulation-to-threshold processor (figure 4.3). This mechanism is required for reproducing the distance and congruence effects described in section 2.2. A

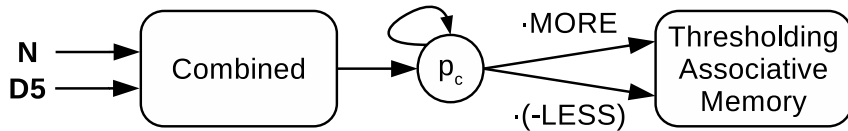


Figure 4.3: Compare processor. *Combined* represents a vector of dimension $2\mathcal{D}$ which is the concatenation of the two input semantic pointers of \mathcal{D} dimensions. The computed comparison is integrated and sent to an autoassociative memory that outputs the answer when a threshold is reached.

combined ensemble represents the concatenation of the semantic pointers corresponding to the two digits to compare (thus it represents $2\mathcal{D}$ -dimensional vectors). The population outputs 1 if the first digit is greater than the second one, -1 if the opposite is true, and 0 if they are equal. This function is implemented using the Computation principle of

the NEF and by randomly selecting vectors corresponding to digit semantic pointers as evaluation points for equation 3.5. The result of the comparison, denoted c^i , is fed to an integrator p_c^i with the following dynamics:

$$\frac{\partial p_c^i}{\partial t} = \frac{c^i}{\tau_c}. \quad (4.2)$$

The time constant τ_c determines the rate of integration and will be optimised to fit human response times. At the beginning of each trial, the integrator is reset manually by sending negative current to its neurons. The semantic pointer corresponding to the fixed reference five, noted **D5**, is given as the second input. Since a positive value indicates that the first digit is greater than five and a negative value indicates the contrary, the output of the integrator is multiplied by pointer **MORE** – **LESS**. It is then sent to a similar associative memory than described above, with a few differences. First, pointers **MORE** and **LESS** are associated with themselves plus **ON**. Therefore, the memory can be considered autoassociative. Second, ensembles receive an inhibitory input θ_{th} . The output of the ensembles is then summed with θ_{th} to compensate for the inhibition. Because ensembles are tuned to only respond to positive values as explained previously, θ_{th} works as a threshold.

4.2 Global workspace

In the global workspace theory, competition for access must satisfy the following conditions:

1. only one coalition can be active in the workspace at a time;
2. the active coalition is maintained in the absence of external inputs to ensure an uninterrupted stream of conscious content;
3. the dynamics of the winning coalition correspond to a nonlinear ignition;
4. in order for a coalition to access the workspace, its activation must be higher than a threshold;
5. processors can form a coalition by proposing a representation collectively, or multiple non-competing representations.

Here I show that winner-take-all networks (WTA), and in particular a neural implementation of a modified leaky competing accumulator model (LCA) (Usher & McClelland, 2001) meet these criteria. The function of WTA networks is to select, among a set of candidate choices, the candidate with the highest activation, and inhibit the other choices. Let ensemble g_j represent the activation of the j -th candidate representation $\mathbf{R}_j \in \mathcal{R}$, \mathcal{R}

containing all representations that can be consciously represented in the workspace. The dynamics of LCA are then given by the following equation:

$$\frac{\partial g_j}{\partial t} = \frac{1}{0.01} \left(g_j^{\leftarrow} - \sum_i g_i \right), g_j \geq 0, \quad (4.3)$$

where g_j^{\leftarrow} is the external input for choice \mathbf{R}_j . In this system, the winning state g_j converges to the value of the largest input g_j^{\leftarrow} , and the losing states converge to zero. However, if input g_j^{\leftarrow} is not maintained, g_j will return to 0. In order to sustain the activation of the winning representation, a feedback term $\Theta \mathcal{H}(g_j)$ is to be added, with \mathcal{H} the Heaviside function. Furthermore, an inhibition weight β_{ij} is introduced to account for the notion of coalition. By default it is set to 1 but it can be 0 when two representations are not meant to compete. Thus, multiple features of a same object can be consciously represented at the same time (e.g. **TASTY** \otimes **APPLE** and **RED** \otimes **APPLE**). The updated equation is:

$$\frac{\partial g_j}{\partial t} = \frac{1}{0.01} \left(g_j^{\leftarrow} + \Theta \mathcal{H}(g_j) - \sum_i \beta_{ij} g_i \right), g_j \geq 0. \quad (4.4)$$

The network shown in figure 5.1 is read out as $g_j^{\rightarrow} = \mathcal{H}(g_j)$. Thus, the output converges to

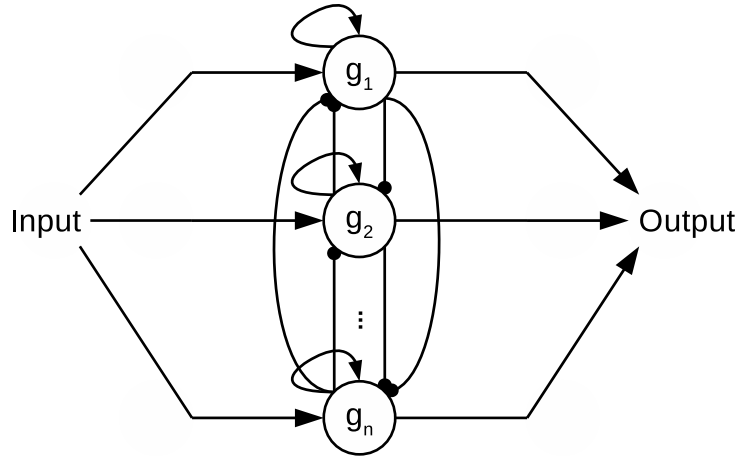


Figure 4.4: Global workspace model. Dynamics are driven by self-amplification and mutual inhibition. Round ends represent inhibitory connections.

one times the winning semantic pointer(s) and is maintained regardless of input g_j^{\leftarrow} . These dynamics correspond to the nonlinear ignition and sustain activation described in global neuronal workspace theory. With the use of the Heaviside function, this also resembles the Independent Accumulator model (IA) from Gosmann et al. (2017), except that no additional layer is introduced. Note that a candidate \mathbf{R}_i must satisfy $g_i^{\leftarrow}(t) > g_j^{\leftarrow}(t) + \Theta$ to become the successor of \mathbf{R}_j . The feedback weight Θ therefore constitutes a threshold for \mathbf{R}_i . For bigger coalitions, the value to exceed is greater as more representations send Θ back to themselves.

4.3 Processors-workspace interactions

It should now become clearer to the reader how specialist processors and a competitive workspace can be implemented in spiking neurons. However, the question of how they interact still needs to be answered. In particular, processors should not contribute equally to the competition, but rather have their contribution weighted by context-dependent attention. Similarly, while information selected in the workspace is made available to all processors, it should only be used by the relevant processors. For example, *Add* should not send candidates to and receive information from the global workspace in the simple and chained-subtraction tasks, but only before the comparison operation of the chained-addition task.

The network proposed by Gosmann (2015) provides a way to multiply two vectors element-wise using the NEF. It can therefore be used to weight the output of the i -th processor to the workspace by an attention control signal a^i , and weight broadcast from the workspace to the processor with another signal b^i . As the candidates are semantic pointers, the output of each processor also needs to be translated into scalar activations. The two weighting networks of the processor implement the following dynamics:

$$\frac{\partial p_j^{i \rightarrow g}}{\partial t} = \frac{1}{0.001} \left(a^i (\mathbf{p}^{i \rightarrow} \cdot \mathbf{R}_j) - p_j^{i \rightarrow g} \right), \quad (4.5)$$

$$\frac{\partial \mathbf{p}^{i \leftarrow g}}{\partial t} = \frac{1}{0.001} \left(-\mathbf{p}^{i \leftarrow g} + \sum_j (b^i g_j^{\rightarrow}) \cdot \mathbf{R}_j \right). \quad (4.6)$$

Small parentheses are meant to emphasise that weighting is done in the activation space of semantic pointers instead of the semantic pointer space for efficiency, as there are less candidate pointers than dimensions in the semantic pointer space. For simplicity, *Visual* does not receive information from the workspace (i.e. $b^V = 0$) and *Motor* does not send information to it (i.e. $a^M = 0$). This does not necessarily hold in real brains where visual processing is modulated by top-down feedback and motor systems can inform higher levels.

The input to the j -th state of the workspace is the sum of contribution of each processor:

$$g_j^{\leftarrow}(t) = \sum_i p_j^{i \rightarrow g}(t). \quad (4.7)$$

Hence, the workspace can select a candidate representation even though that representation was not identified as the most relevant one by any processor. In that sense, processors can form a coalition to bring the same representation to consciousness. As explained in the previous part, processors can also propose multiple non-competing representations (e.g. describing different features of an object) to the workspace. As the

study of Sackur & Dehaene (2009) suggests, there is also direct inter-processor communication (see crosstalk connection in figure 4.1). This can also be considered a form of collaboration. The input to the i -th processor, denoted by $\mathbf{p}^{i\leftarrow}$, can thus be modelled as the combination of information coming from processors $j \neq i$ weighted by crosstalk weights $\omega_{j \rightarrow i}$, sensory input \mathbf{s} , and broadcast weighted by b^i :

$$\mathbf{p}^{i\leftarrow}(t) = \mathbf{s}(t) + \mathbf{p}^{i\leftarrow g}(t) + \sum_{j \neq i} \omega_{j \rightarrow i} \mathbf{p}^{j \rightarrow}(t). \quad (4.8)$$

In summary, although the notion of coalition is not clearly defined in the literature, the model presented here accounts for its broad description and lays the ground for its mathematical formalism.

4.4 Router

In the previous part, I introduced attention and broadcast control signals. They must be set according to the task to be performed and the task state of progress. This is done by *Routing Controller*. As described in section 3.2, cortico-basal ganglia-thalamo-cortical loops can implement cognitive control. More precisely, they can recognise the satisfaction of a production rule condition in cortical areas, and execute a corresponding action. Here I propose that control signals \mathbf{a} and \mathbf{b} are set according to what is being represented in the workspace, in the processors, and in working memory components that store the identity of the task and the production rule that was previously executed.

4.4.1 Working memory

From the beginning of the trial, *Task* is given the semantic pointer corresponding to the task: SIMPLE, CHAINED_ADD or CHAINED_SUB. Note that this process should not be done manually, but rather modelled as the perception of the task from the screen and the routing of that information to *Task*. The behaviour of the model during trials is not influenced by this limitation, as participants were informed about the task to perform before a trial block started. However, it leaves room for future extension of the model.

Previous Routing is a working memory component that keeps track of the last command executed by the router. Its dynamics are essentially the same as those of the comparison integrator (eq. 4.2), but in the \mathcal{D} -dimensional space where semantic pointers lie. Its integration time constant is denoted by τ_r .

4.4.2 Production rules

Routing Controller is composed of a set of cortico-basal ganglia-thalamo-cortical loops implementing production rules. Lateral inhibition ensures that rules are mutually exclusive (figure 3.1). As loops work best with positive utilities, the utility of each rule gets a constant bias 0.5. First, a “Thresholding” rule with constant utility 0.6 (outvalues the bias by 0.1) allows the controller to be inactive when other rules have a lower utility:

$$\begin{array}{ll} \text{if} & 0.6 \\ \text{is maximal, then} & \emptyset. \end{array}$$

The “Get *Visual*” rule is executed when the fixation cross preceding the presentation of N is represented in *Visual* or the workspace. It increases the attention control signal of *Visual* so that visual content enters the workspace. The constant $\alpha_a = 20$ ensures the output of processors is high enough to pass the threshold Θ and be granted access to the workspace. In addition, like other rules except “Thresholding”, the identity of the rule is sent to *Previous Routing* (noted *prev*):

$$\begin{array}{ll} \text{if} & 0.5 + \mathbf{p}^{V \rightarrow} \cdot \mathbf{FIXATE} + g_{\mathbf{FIXATE}}^{\rightarrow} \\ \text{is maximal, then} & \alpha_a \gg a^V, \\ & \mathbf{GET} \otimes \mathbf{V} \gg \mathbf{prev}. \end{array}$$

“Set *Add*” is executed when “Get *Visual*” has just been executed and the model needs to performed the chained-addition task. It is selected until preconscious content is available in *Add* (indicated by the activation of **ON**). The broadcast control signal of *Add* is then increased:

$$\begin{array}{ll} \text{if} & 0.5 - g_{\mathbf{FIXATE}}^{\rightarrow} + \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{ADD} \times (0.5 - \mathbf{p}^{\mathbf{ADD} \rightarrow} \cdot \mathbf{ON}) \\ & + \mathbf{prev} \cdot \mathbf{GET} \otimes \mathbf{V} \times \mathbf{task} \cdot \mathbf{CHAINED_ADD} \\ \text{is maximal, then} & 1 \gg b^{\mathbf{ADD}}, \\ & \mathbf{SET} \otimes \mathbf{ADD} \gg \mathbf{prev}. \end{array}$$

“Set *Sub*” is similar to “Set *Add*”, except that it is executed in the chained-subtraction task:

$$\begin{aligned}
& \text{if} && 0.5 - g_{\text{FIXATE}}^{\vec{}} + \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{SUB} \times (0.5 - p^{\text{SUB} \rightarrow} \cdot \mathbf{ON}) \\
& && + \mathbf{prev} \cdot \mathbf{GET} \otimes \mathbf{V} \times \mathbf{task} \cdot \mathbf{CHAINED_SUB} \\
\text{is maximal, then} && 1 \gg b^{\text{SUB}}, \\
& && \mathbf{SET} \otimes \mathbf{SUB} \gg \mathbf{prev}.
\end{aligned}$$

“Set *Compare*” is slightly more complex because it can be executed in different conditions according to the task: when *Visual* content has been called to the workspace in the simple task, when *Add* content has been called in the chained-addition task, or when *Subtract* content has been called in the chained-subtraction task.

$$\begin{aligned}
& \text{If} && 0.5 - g_{\text{FIXATE}}^{\vec{}} + \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{COM} \times (0.5 - p^{\text{COM} \rightarrow} \cdot \mathbf{ON}) \\
& && + \mathbf{prev} \cdot \mathbf{GET} \otimes \mathbf{V} \times \mathbf{task} \cdot \mathbf{SIMPLE} \\
& && + \mathbf{prev} \cdot \mathbf{GET} \otimes \mathbf{ADD} \times \mathbf{task} \cdot \mathbf{CHAINED_ADD} \\
& && + \mathbf{prev} \cdot \mathbf{GET} \otimes \mathbf{SUB} \times \mathbf{task} \cdot \mathbf{CHAINED_SUB} \\
\text{is maximal, then} && 1 \gg b^{\text{COM}}, \\
& && \mathbf{SET} \otimes \mathbf{COM} \gg \mathbf{prev}.
\end{aligned}$$

When “Set *Add*” has just been executed and some representation is preconscious in *Add*, that representation is brought into the workspace by executing “Get *Add*”, which increases the corresponding attention control signal. Typically, the condition is satisfied when the result of $N \oplus 2$ is preconscious after broadcasting N from the workspace to *Add*:

$$\begin{aligned}
& \text{if} && 0.5 + \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{ADD} \times p^{\text{ADD} \rightarrow} \cdot \mathbf{ON} \\
\text{is maximal, then} && \alpha_a \gg a^{\text{ADD}}, \\
& && \mathbf{GET} \otimes \mathbf{ADD} \gg \mathbf{prev}.
\end{aligned}$$

Rules “Get *Subtract*” and “Get *Compare*” are similar to “Get *Add*”:

$$\begin{aligned}
& \text{if} && 0.5 + \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{SUB} \times p^{\text{SUB} \rightarrow} \cdot \mathbf{ON} \\
\text{is maximal, then} && \alpha_a \gg a^{\text{SUB}}, \\
& && \mathbf{GET} \otimes \mathbf{SUB} \gg \mathbf{prev}.
\end{aligned}$$

If $0.5 + \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{COM} \times \mathbf{p}^{COM \rightarrow} \cdot \mathbf{ON}$
is maximal, then $\alpha_a \gg a^{COM}$,
 $\mathbf{GET} \otimes \mathbf{COM} \gg \mathbf{prev}$.

Finally, once the result of the comparison operation has been computed and sent to the workspace, it can be reported by executing “Set *Motor*”. Note that the utility of that rule gets a positive feedback of 0.5 in order to maintain the motor output and stop the execution of other rules until the fixation cross is shown again to prepare the next trial:

if $0.5 + \mathbf{prev} \cdot \mathbf{GET} \otimes \mathbf{COM} + 0.5 \times \mathbf{prev} \cdot \mathbf{SET} \otimes \mathbf{M}$
is maximal, then $\alpha_a \gg b^M$,
 $\mathbf{SET} \otimes \mathbf{M} \gg \mathbf{prev}$.

Chapter 5

Results

5.1 Model behaviour

In this section, the behaviour of the model performing the three tasks ($N \leq 5?$, $N \ominus 2 \leq 5?$, $N \oplus 2 \leq 5?$) with $N = 2$ is shown. Parameters are set to $\omega_{V \rightarrow COM} = 0.5$, $\tau_r = 0.05$, $\tau_c = 0.05$ and $\mathcal{C} = 0.25$ (this parameter is described in the next section) for illustration purposes. The dimensionality of semantic pointers was set to $\mathcal{D} = 96$.

As shown in figure 5.1, the input stimulus is presented during 29 ms following a fix-

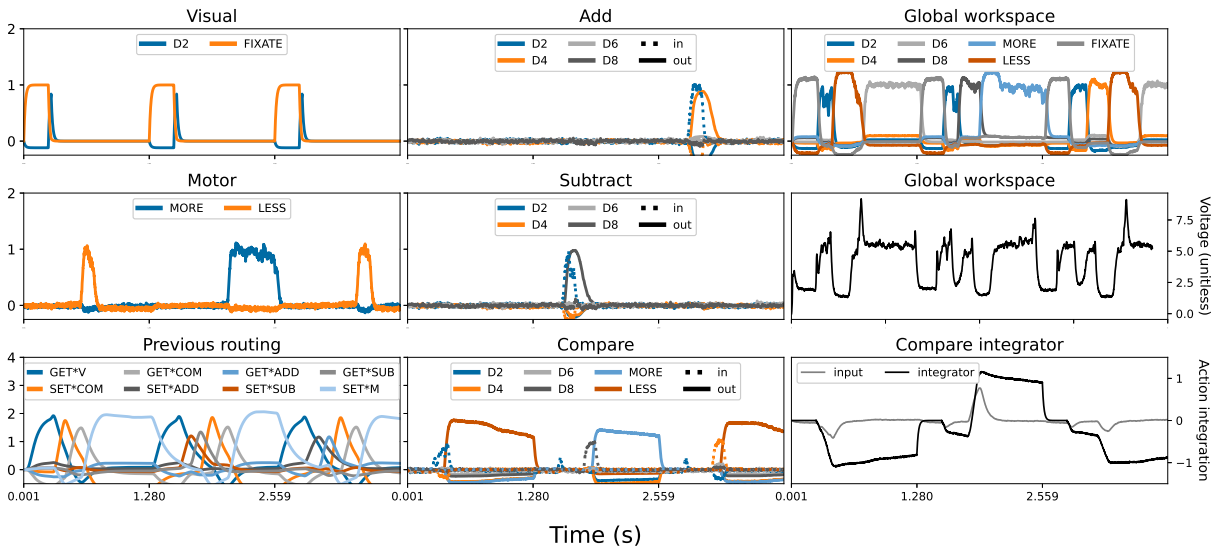


Figure 5.1: Behaviour of the model successively performing the simple, chained-subtraction and chained-addition tasks. Unless otherwise specified, the vertical axes represent similarity with semantic pointers (computed as the dot product between pointer vectors and activity). Dotted lines represent processors input. The average internal voltage of workspace neurons is shown in the middle right panel. The bottom right panel shows evidence accumulation in *Compare* towards the input being greater (positive values) or less than five (negative values).

ation cross lasting 250 ms. The model then correctly sends information to the relevant

processors and outputs the right answer. Because of the crosstalk connection between *Visual* and *Compare*, the comparison starts with N and biases evidence accumulation towards the wrong answer in the chained-subtraction task (the trial is incongruent). The content of *Previous Routing* and the workspace provide clear indications about the routing rules that are being executed and the concept that is being consciously represented, respectively. When a new representation is granted access to the workspace, the voltage exhibits a positive wave resembling P3 components. It is also observed in a more controlled setting (figure 5.2).

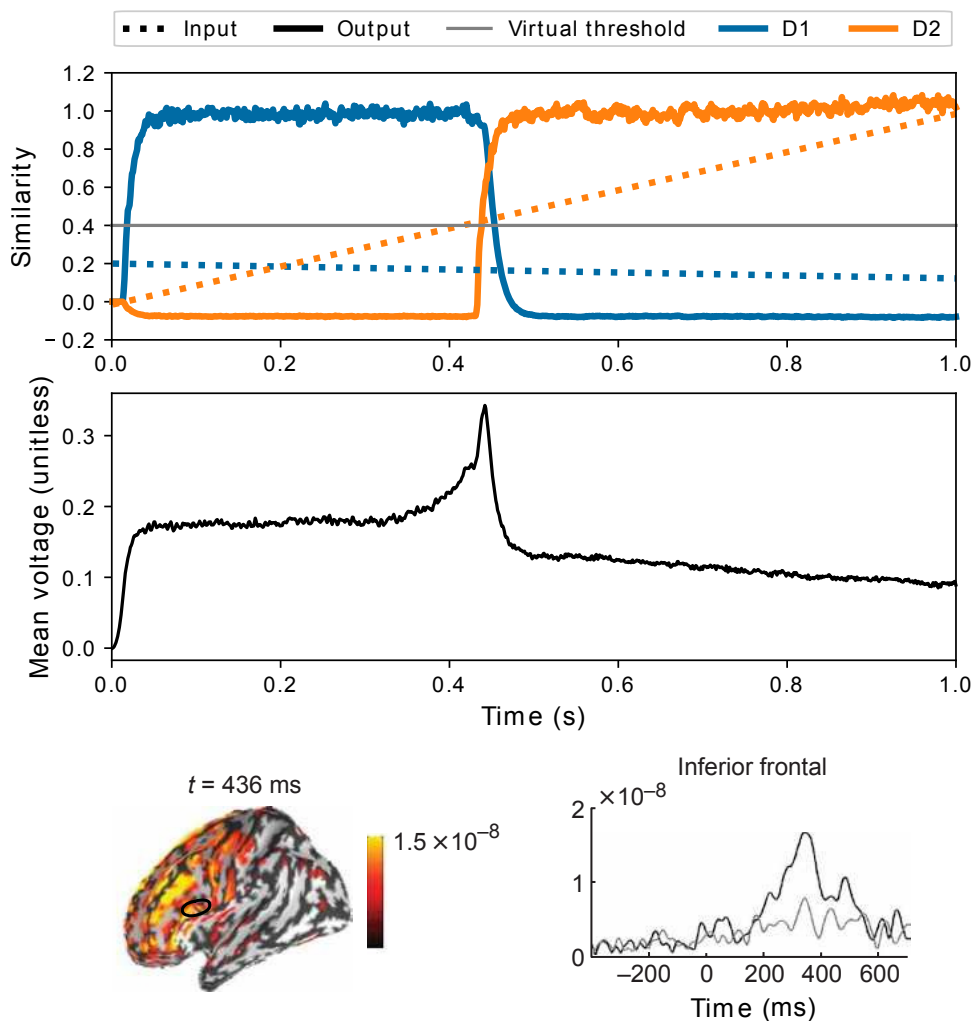


Figure 5.2: Positive wave indexing competition for access consciousness in the model (top two panels) and the human brain (lower two panels; adapted from Sergent et al., 2005). (Top) Pointer **D1** is presented with a constant value of 0.2 and **D2** increases linearly with time. The value $0.2 + \Theta = 0.4$ that **D2** has to exceed is shown in gray. (Middle) When **D2** gets selected after passing the threshold, the internal voltage of workspace neurons exhibits a positive wave. (Bottom) A similar wave (P3) can be observed with EEG in the human frontal cortex when a stimulus is seen (black), but not when it is subliminal (gray).

In section 4.2, I explained how the workspace is able to integrate multiple non-

competing representations. This is illustrated in figure 5.3 with a toy example. In particular, integration is both synchronic and diachronic. The workspace indeed integrates information from the past (**REMEMBER**), the present sensory information (**HEAR** and **SEE**), and planned actions (**FEED**) simultaneously, as inspired by figure 2.2. Notice how the threshold increased in comparison with the previous example, as two representations self-amplify rather than one.

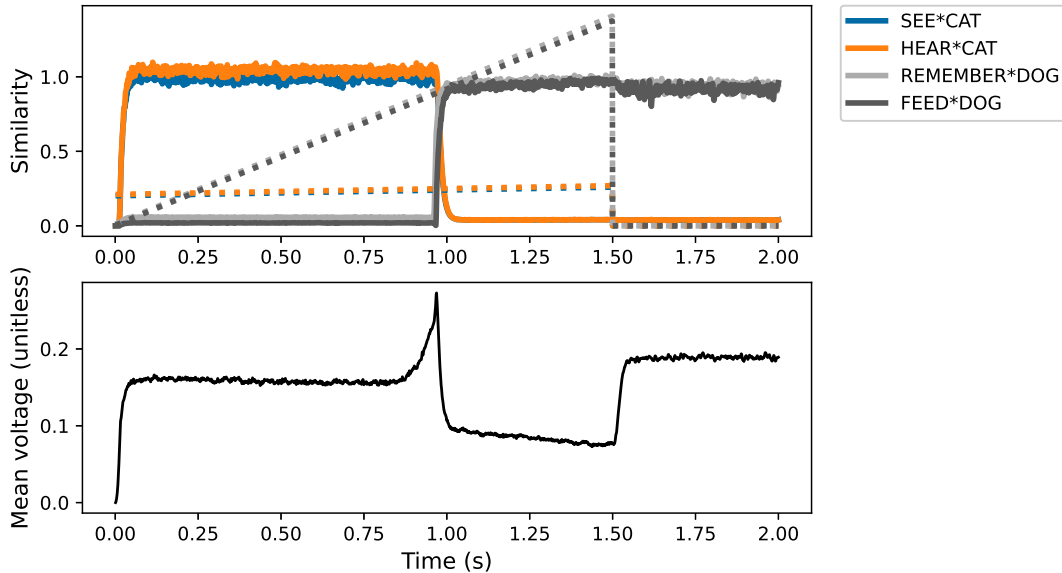


Figure 5.3: Example of coalition formation. Pointer **SEE** \otimes **CAT** + **HEAR** \otimes **CAT** is presented with a value of 0.2 and **REMEMBER** \otimes **DOG** + **FEED** \otimes **DOG** increases linearly with time. The value to exceed is now $2 \times 0.2 + 2\Theta = 0.8$. After 1.5 s, the input is stopped to show that the network is capable of maintaining multiple active representations.

5.2 Empirical fitness

5.2.1 Parameters

Using the approach described in section 3.3, the fitness evaluation function is optimised with respect to a selection of parameters that I will describe now. The number of neurons in the *combined* ensemble is set to $100\mathcal{DC}$, with \mathcal{D} the number of semantic pointer dimensions and \mathcal{C} a scaling factor. This value appears to scale the distance effect associated with the comparison operation (figure 5.4). Time constant τ_c determines the rate of integration of the comparison integrator, and therefore the time for *Compare* to decide whether the input is greater or less than five. τ_r drives information integration in *Previous Routing*. Thus, it influences the speed at which *Routing Controller* selects new rules. Crosstalk connection $\omega_{V \rightarrow COM}$ is expected to cause the comparison operation to

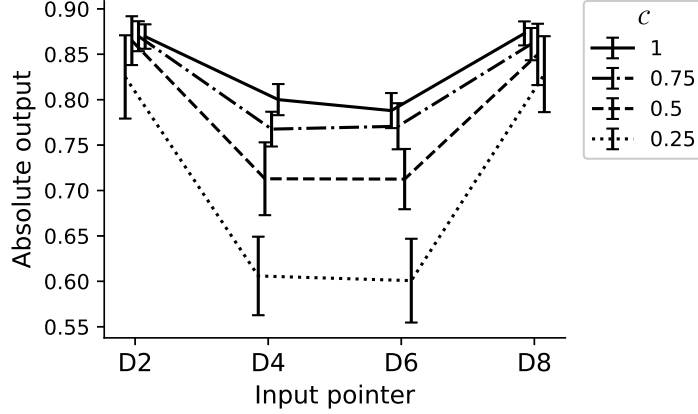


Figure 5.4: Output of the *combined* ensemble which compares two digit semantic pointers ($\mathcal{D} = 96$) as a function of the second input and the number of neurons. The first input is the reference **D5**. The distance effect magnitude decreases as the number of neurons in the ensemble increases. Note that the absolute output is shown, yet the output is negative for inputs **D2** and **D4**, and positive for **D6** and **D8**. Error bars represent plus and minus one standard deviation from the mean computed with 30 seeds.

start with N instead of waiting for the arithmetical operation to be performed in chained trials. Therefore, this parameter likely modulates the congruence effect described in section 2.2. Finally, the time taken by unmodelled sensory and motor processing, denoted by θ_s , is optimised within each Bayesian optimisation call as explained in section 3.3.

5.2.2 Optimisation

Parameters are first optimised to fit response times only ($\alpha_{RT} = 1$ and $\alpha_{ER} = 0$). The dimensionality of semantic pointers is set to $\mathcal{D} = 96$. The value of the evaluation function \mathcal{L} converges towards a minimal value of 10.55 (figure 5.5). Figure 5.6 shows how function

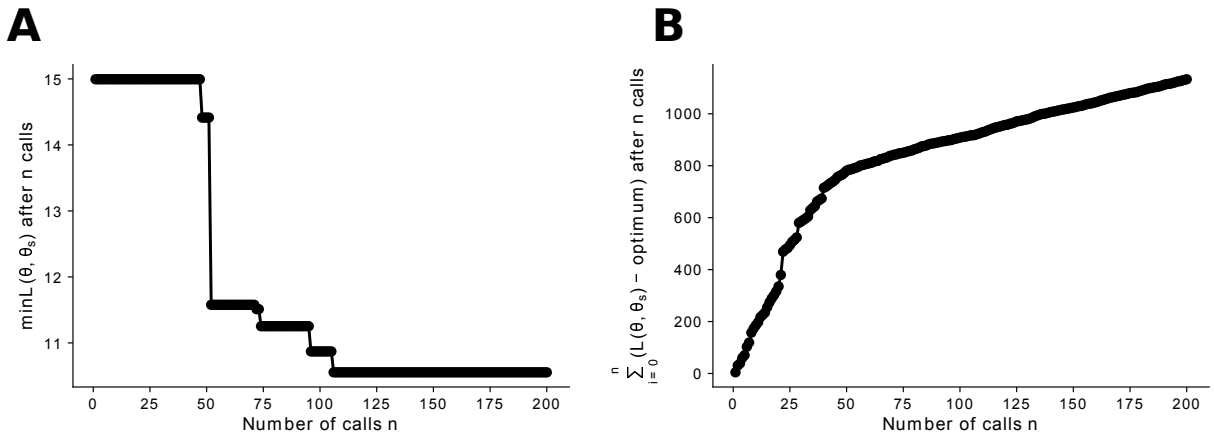


Figure 5.5: (A) Analysis of convergence computed as the empirical optimum of $\mathcal{L}(\theta, \theta_s)$ found after n calls. (B) Cumulative regret computed as the sum of function evaluations minus the empirical optimum found over the whole optimisation procedure.

values vary with respect to individual and pairs of parameters. The best fit is obtained with the lowest time constants, indicating that humans perform the tasks in a small amount of time. As it also appears, \mathcal{C} should not be very low, although the best call had $\mathcal{C} = 0.1$. Interestingly, the function is low in relatively broad regions of the parameter space.

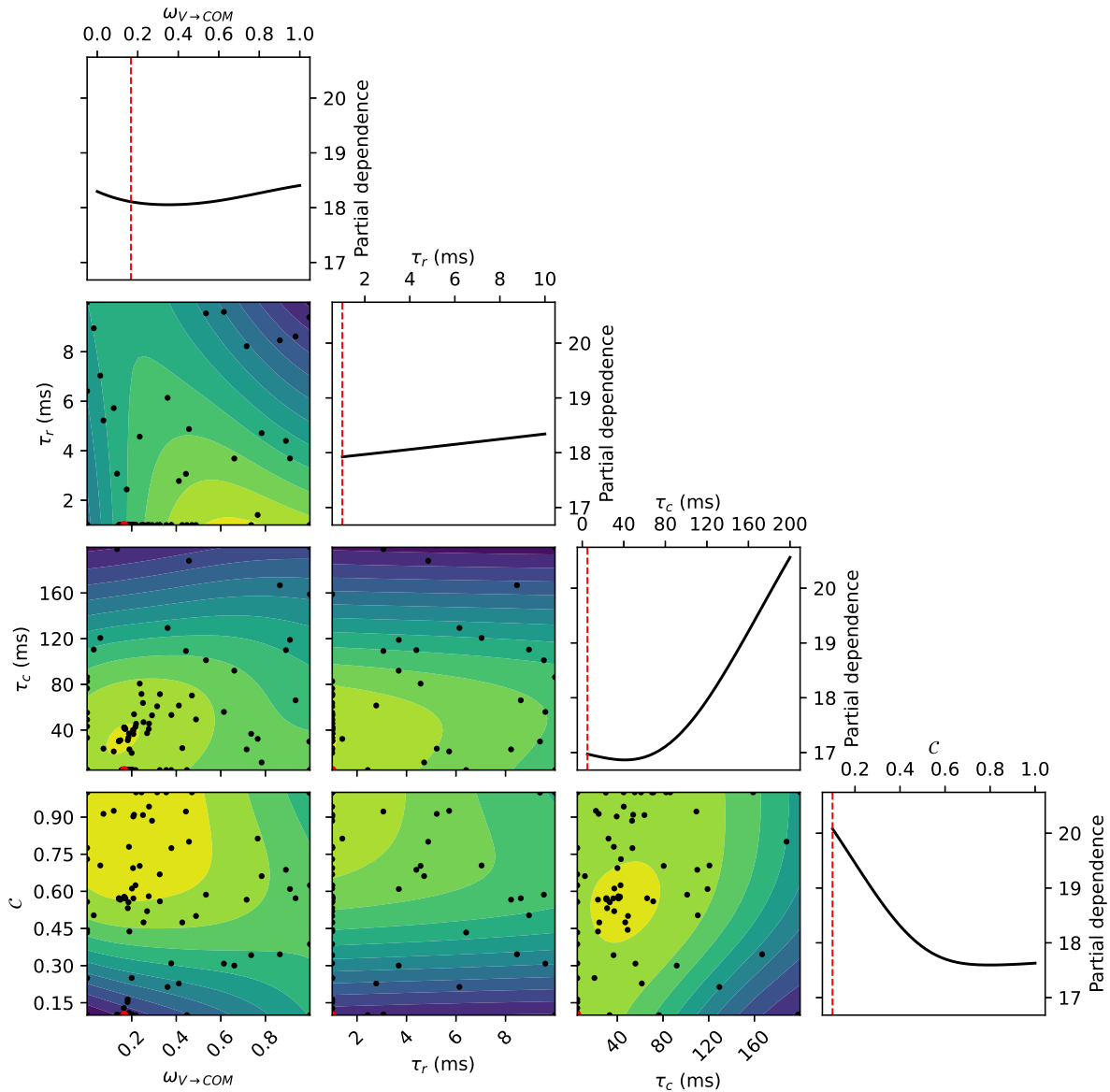


Figure 5.6: Partial dependence of parameters for fitting human response times ($\alpha_{RT} = 1$ and $\alpha_{ER} = 0$). Dots indicate the points at which the function was evaluated. The found minimum is shown with red dots and red dashed lines. The function is lowest in yellow areas and highest in blue areas.

Error rates and median correct response times of the model with the best parameters found are shown in figure 5.7. For better statistical power, more participants and conditions were simulated. Instead of performing each condition twice like in the opti-

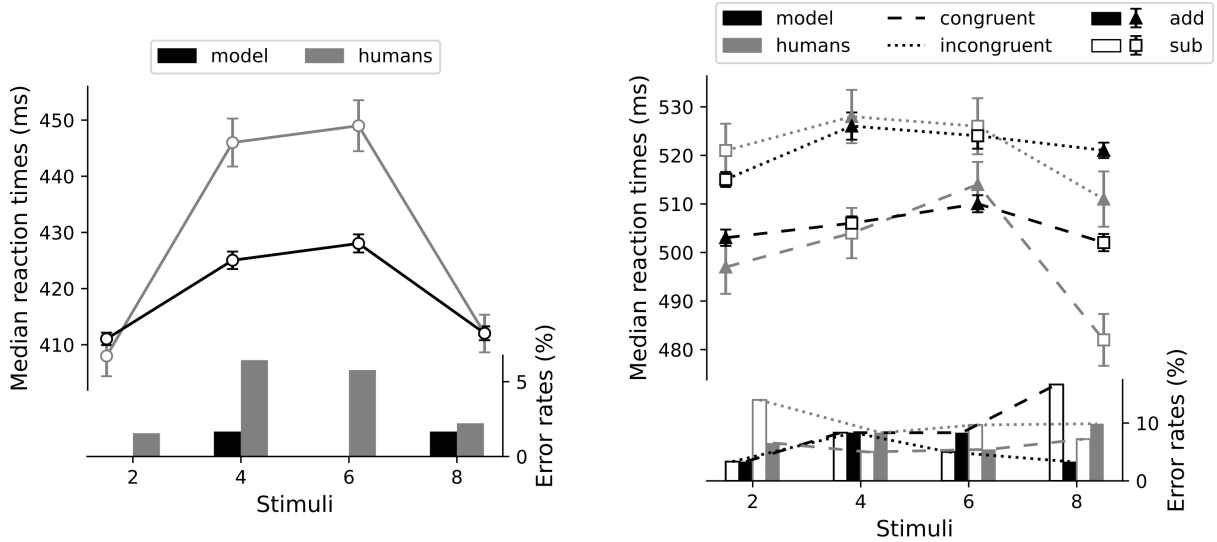


Figure 5.7: Comparison of the behaviour of humans and the model with the best parameters found ($\omega_{V \rightarrow COM} = 0.166$, $\tau_r = 0.001$, $\tau_c = 0.005$, $\mathcal{C} = 0.1$, $\tau_s = 257$ ms) by response times optimisation. Simple trials are on the left and chained trials on the right. On top are median response times with error bars representing plus and minus one standard error of the mean. At the bottom, the error rates are shown. For chained trials, congruent and incongruent conditions are emphasised by dashed and dotted lines respectively.

misation procedure, the model performed each condition ten times. Moreover, 30 “participants” were simulated instead of 10. Three of them were excluded from the analyses as they did not answer correctly in any trial of the chained-subtraction task. Median response time was 94 ms higher in the chained tasks than in the simple task (difference was 85 ms in humans). In the simple task, stimuli closest to five (4 and 6) led to a slower response than furthest stimuli (2 and 8) by 15 ms (36 ms in humans). Finally, in chained trials, median response time was lower in congruent trials (stimulus and image fell on the same side of five) than in incongruent trials (stimulus and image did not fall on the same side) by 14 ms (20 ms in humans). To check the significance of these effects, ANOVAs were performed on response times¹. The first one took stimuli (2, 4, 6, 8) and operations ($N \leq 5?$, $N \oplus 2 \leq 5?$, $N \ominus 2 \leq 5?$) as factors. The effect of operation was significant ($F(2) = 16922.572$, $p < .001$). A significant effect of stimuli was also found ($F(3) = 64.248$, $p < .001$). The operation \times stimuli interaction was significant too ($F(6) = 72.356$, $p < .001$). To test whether human and model behaviours could be distinguished, a new ANOVA was performed with humanity (true or false) as an additional factor. The operation \times stimuli \times human interaction was not significant ($F(6) = 1.394$, $p = .213$) which provides strong evidence in favour of the model validity.

Another ANOVA was performed on simple trials to check whether the distance effect

¹Note that in Sackur & Dehaene (2009), ANOVAs were performed on median correct RTs with subject as random factor. However, since the data likely contains some mistakes (see the footnote in section 3.3), results obtained with this approach were not consistent with those reported in the original study.

was the same in humans and the model. The factors were stimuli and humanity. The interaction was significant ($F(3) = 5.967, p < .001$). In fact, figure 5.7 shows that the distance effect is clearly smaller in the model. A follow-up ANOVA was performed to analyse the effect of stimulus-image congruency in the model and in humans (chained trials only). It took operations, congruency and humanity as factors. The interaction between operations and congruency was significant ($F(1) = 8.775, p < .005$). Yet, congruence effects of the model and humans could not be distinguished, as the operation \times congruency \times humanity interaction was not significant ($F(1) = 2.222, p = .136$). Finally, in order to investigate the presence of potential stimulus and image distance effects (with respect to five) in chained trials, an ANOVA with stimulus distance (close: stimulus is 4 or 6, far: 2 or 8) and image distance (close: image is 4 or 6, far: 2 or 8) as factors was performed. None of the main effects were significant ($ps > .2$). The interaction was significant ($F(1) = 220.752, p < .001$), but reflects only the stimulus-image congruence effect. Similar results were reported in the original study.

Clearly, the model does not reproduce the pattern of errors exhibited by humans (figure 5.7). More precisely, the distance effect in simple trials and stimulus-image congruence effects in chained trials are not reproduced. In fact, model performance was poorer in congruent trials. To ensure this was not caused by the evaluation function (which only took response times into consideration), a new optimisation was performed with $\alpha_{RT} = 1$ and $\alpha_{ER} = 1$. Moreover, the dimensionality of semantic pointers (\mathcal{D}), which is typically correlated with performance, was used as an additional parameter to optimise. Although distance and congruent effects were preserved in response times, this approach did not result in improved fitness of error rates (figure 5.8).

5.3 Performance on long tasks

As a final evaluation of the model, I investigate whether it is capable of chaining many operations. The task is to increment one M times starting from zero and thus report M . In order to do that, $M + 1$ digit semantic pointers are generated ($\mathbf{D0}, \dots, \mathbf{D}_i, \dots, \mathbf{D}_M$). *Visual* is given $\mathbf{D0}$ as input and the model has to chain information through M *Add* processors responsible for adding 1 to their input. To prevent the model from taking shortcuts, the i -th *Add* processor can only recognise semantic pointer \mathbf{D}_i and output \mathbf{D}_{i+1} . A trial is successfully completed if *Motor* receives \mathbf{D}_{i+1} with an activation of at least 0.5 at some time within $0.25 \times (M + 2)$ seconds. The results are shown in figure 5.9. As might be expected, accuracy increases with semantic pointer dimensionality \mathcal{D} and decreases with the number of additions to perform. Yet, the model is able to perform 50 additions and report $\mathbf{D50}$ in 82% of trials with $\mathcal{D} = 128$. Hence, the model can chain many more operations than required in the tasks for which it has been designed.

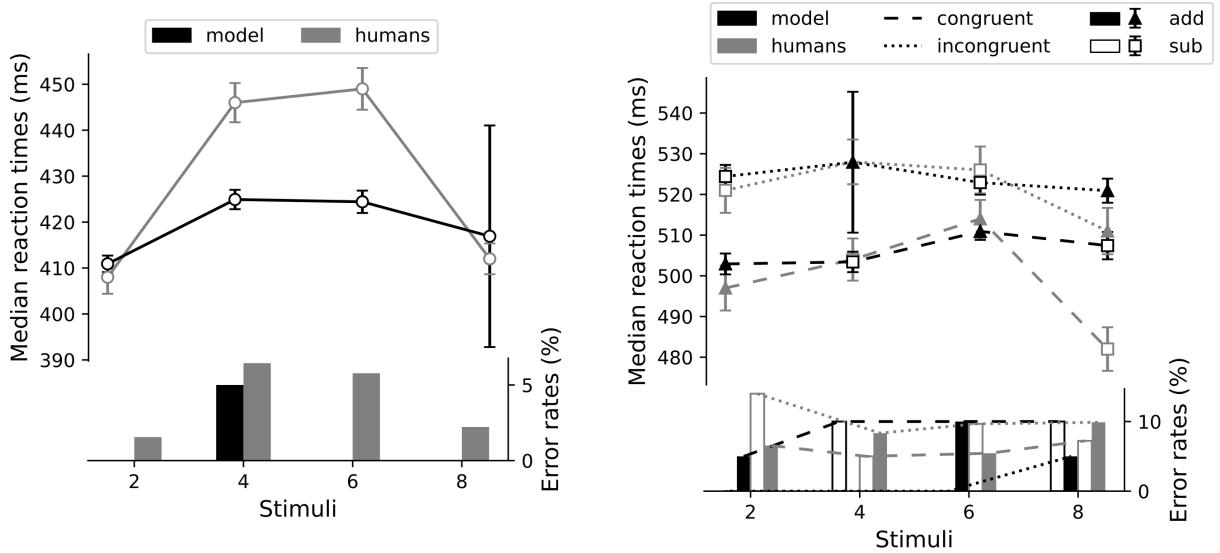


Figure 5.8: Comparison of the behaviour of humans and the model with the best parameters found ($\omega_{V \rightarrow COM} = 0.227$, $\tau_r = 0.001$, $\tau_c = 0.049$, $\mathcal{C} = 1$, $\mathcal{D} = 160$, $\tau_s = 209$ ms) by optimising both response times and error rates.

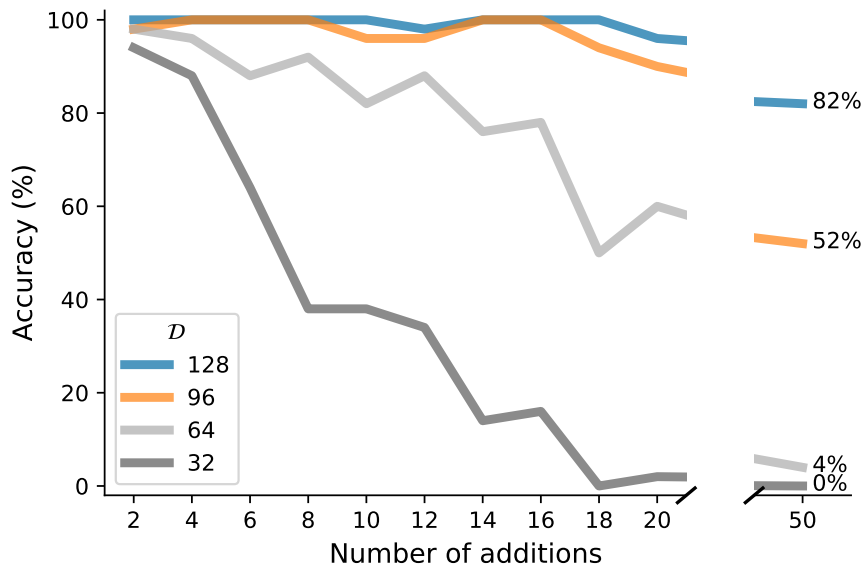


Figure 5.9: Accuracy as a function of the number of additions to perform (M) and the number of dimensions. The accuracy in each condition is evaluated on 50 trials with different seeds for the random generators. Percentages on the right indicate performance for each dimensionality at $M = 50$.

With the absence of crosstalk, the execution of the task is purely serial. Therefore, there should exist a probability of success p_D for performing one addition corresponding to dimensionality D . Thus, the probability of successfully performing M additions is $p_{D,M} = (p_D)^M$. Probability p_D can be estimated empirically from the accuracies of

simulations:

$$p_{\mathcal{D}} = \frac{1}{11} \sum_{i=1}^{11} (p_{\mathcal{D},M_i})^{1/M_i}. \quad (5.1)$$

The value 11 corresponds to the number of M values used in the simulations ($M_i = 2, 4, 6, \dots, 50$). For better predictions, conditions without any successful trial are excluded from the computation (e.g. $M = 18, \mathcal{D} = 32$). This simple approach correctly predicts the performance of the model as shown in figure 5.10 ($r = .962, p < .001$). More precisely, correlations were significant for $\mathcal{D} = 32$ ($r = .975, p < .001$), $\mathcal{D} = 64$ ($r = .961, p < .001$), $\mathcal{D} = 96$ ($r = .928, p < .001$), and $\mathcal{D} = 128$ ($r = .925, p < .001$).

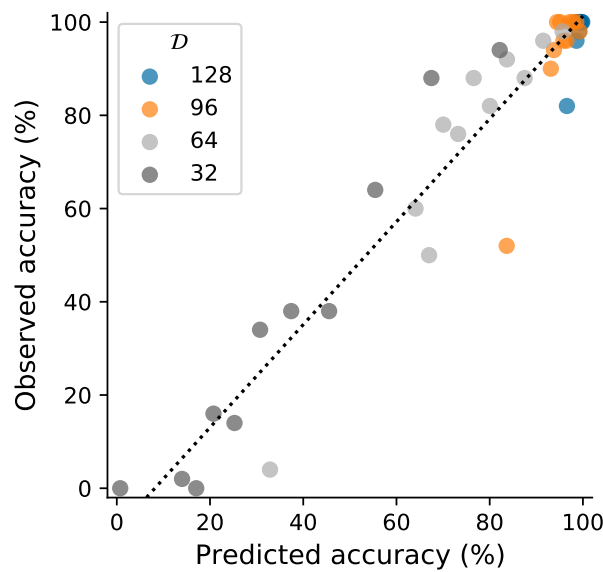


Figure 5.10: Prediction of accuracy. Each point corresponds to a condition from the previous figure. The prediction is made according to the following formula: $p_{\mathcal{D},M} = (p_{\mathcal{D}})^M$. The dotted line represents the regression line.

Chapter 6

Discussion

The model I propose here reproduces the cognitive functions associated with global neuronal workspace theory (section 4). Processors are composed of virtually any set of processes and compete to bring their output into the workspace. As illustrated in figure 5.3, only one coalition can be represented in the workspace at a time, and is maintained until a new coalition outvalues it by a certain threshold. Conscious content is broadcast to the input of all processors. Gating mechanisms modulate how much information is sent from each processor, and how much each processor can use conscious content for further processing. Crucially, gating is controlled according to the task to be performed and its current progress. Unlike Zylberberg et al. (2010)'s model, mine can perform many steps including information routing between any processors provided that the number of dimensions in the representational space is sufficiently large (section 5.3). Furthermore, the number of neurons in the workspace does not scale with the number of processors and symbols, but only with the number of candidate conscious representations.

The model is anatomically and physiologically plausible. Neural constants in various regions are consistent with their biological counterparts. However, workspace neurons are often said to be connected through excitatory connections (Dehaene et al., 1998). Although winner-take-all computation is implemented with both inhibitory and excitatory synapses in the model (a default mode of the NEF), constraints on the connections, along with more biological details, could be implemented using an extension of the NEF (Stöckel & Eliasmith, 2019) at a potential computational cost.

To my knowledge, it is the first neural model accounting for the formation of coalitions. In particular, the global workspace not only performs sensory integration, but also integrates information of different time windows (past and future; figure 5.3). This is made easier by the use of SPA, as the binding operation of vector symbolic architectures have been shown to solve the binding problem (Gayler, 2004). The model makes an interesting prediction in this regard. As the size of the conscious coalition increases, the threshold for conscious access also increases. This was predicted when creating the model

(section 4.2) and confirmed in simulations (figure 5.3). The prediction could be tested empirically in a multisensory minimal contrast paradigm. Unfortunately, most studies related to access consciousness focus on a single sensory modality (typically vision).

In the framework proposed by Zylberberg et al. (2011, not to be confused with the neural model mentioned above), each execution of a production rule is indexed by a P3 component. In my model, conscious content is not directly linked to the production system, and P3 waves rather index conscious content (figures 5.1 and 5.2), in accordance with global neuronal workspace theory (Mashour et al., 2020). A processing step is characterised by the execution of a “Set” and a “Get” rule (defined in section 4.4). Therefore, with the same rules, the framework predicts that the execution of one step is indexed by two P3 waves, while my model predicts the appearance of only one wave (when the result enters the workspace after executing “Get”). This could also be tested empirically.

Most importantly, the model reproduces the patterns of response times reported by Sackur & Dehaene (2009), as illustrated in figure 5.7. In some analyses, the behaviour of the model is indistinguishable from humans (section 5.2.2). Namely, the congruence effect was statistically equivalent to that of humans. Moreover, a smaller but significant distance effect was present. The model also provides a computational account as to how the congruence effect levels off stimulus and image distance effects in chained trials. However, the model did not reproduce human error rates (figure 5.8). As the comparison integrator was reset manually, it lacks a control mechanism and is therefore an incomplete model of its biological counterpart. A better model may be more prone to mistakes and reproduce human errors, but it poses the question as to how the control signal should be generated. It might be computed locally, for example according to the amount of input semantic pointers, or be generated by the global workspace or the router. Another possibility is that the integrator is “leaky” and goes back to 0 at a constant rate. Although these approaches have been considered for this work, the choice for manual reset was governed by the lack of evidence in support of alternative approaches but also and more importantly because a perfect integrator is much more stable and predictable. In fact, the choice of a stable perfect integrator may also have been the cause of lower variability in response times (not directly shown, but suggested by figures 5.7 and 5.8).

To summarise, the proposed functional neural model is robust as well as psychologically and biologically plausible. It bridges the gap between high-level functions and low-level specifications of the global workspace organisation. Moreover, it makes testable predictions at both levels. Hence, it provides a suitable framework for further developing and evaluating the validity of global workspace theory in relation with complex serial processing.

Bibliography

- Baars, B. J. (1993). *A cognitive theory of consciousness*. Cambridge University Press.
- Baars, B. J., & Franklin, S. (2009). Consciousness is computational: The lida model of global workspace theory. *International Journal of Machine Consciousness*, 1(01), 23–32.
- Baars, B. J., Franklin, S., & Ramsøy, T. Z. (2013). Global workspace dynamics: cortical “binding and propagation” enables conscious contents. *Frontiers in psychology*, 4, 200.
- Bekolay, T., Bergstra, J., Hunsberger, E., DeWolf, T., Stewart, T. C., Rasmussen, D., Choo, X., Voelker, A., & Eliasmith, C. (2014). Nengo: a python tool for building large-scale functional brain models. *Frontiers in neuroinformatics*, 7, 48.
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, 16(5), 1190–1208.
- Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of consciousness studies*, 2(3), 200–219.
- Connor, D., & Shanahan, M. (2010). A computational model of a global neuronal workspace with stochastic connections. *Neural Networks*, 23(10), 1139–1154.
- Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. In *Seminars in the Neurosciences*, vol. 2, (pp. 263–275). Saunders Scientific Publications.
- Crick, F., & Koch, C. (2003). A framework for consciousness. *Nature neuroscience*, 6(2), 119–126.
- Dehaene, S. (2014). *Consciousness and the brain: Deciphering how the brain codes our thoughts*. Penguin.
- Dehaene, S., & Changeux, J.-P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS Biol*, 3(5), e141.
- Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2), 200–227.

- Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in cognitive sciences*, *10*(5), 204–211.
- Dehaene, S., Charles, L., King, J.-R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current opinion in neurobiology*, *25*, 76–84.
- Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, *95*(24), 14529–14534.
- Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition*, *79*(1-2), 1–37.
- Dehaene, S., Naccache, L., Cohen, L., Le Bihan, D., Mangin, J.-F., Poline, J.-B., & Rivière, D. (2001). Cerebral mechanisms of word masking and unconscious repetition priming. *Nature neuroscience*, *4*(7), 752–758.
- Dehaene, S., Sergent, C., & Changeux, J.-P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences*, *100*(14), 8520–8525.
- Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biol*, *5*(10), e260.
- Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. Oxford University Press.
- Eliasmith, C., & Anderson, C. H. (2003). *Neural engineering: Computation, representation, and dynamics in neurobiological systems*. MIT press.
- Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *science*, *338*(6111), 1202–1205.
- Fan, Z., Muthukumaraswamy, S. D., Singh, K. D., & Shapiro, K. (2012a). The role of sustained posterior brain activity in the serial chaining of two cognitive operations: A meg study. *Psychophysiology*, *49*(8), 1133–1144.
- Fan, Z., Singh, K., Muthukumaraswamy, S., Sigman, M., Dehaene, S., & Shapiro, K. (2012b). The cost of serially chaining two cognitive operations. *Psychological research*, *76*(5), 566–578.
- Fodor, J. A. (1983). *The modularity of mind*. MIT press.

- Fountas, Z., Gamez, D., & Fidjeland, A. K. (2011). A neuronal global workspace for human-like control of a computer game character. In *2011 IEEE Conference on Computational Intelligence and Games (CIG'11)*, (pp. 350–357). IEEE.
- Gaillard, R., Dehaene, S., Adam, C., Clémenceau, S., Hasboun, D., Baulac, M., Cohen, L., & Naccache, L. (2009). Converging intracranial markers of conscious access. *PLoS Biol*, *7*(3), e1000061.
- Gayler, R. W. (2004). Vector symbolic architectures answer jackendoff’s challenges for cognitive neuroscience. *arXiv preprint cs/0412059*.
- Gosmann, J. (2015). Precise multiplications with the NEF. Tech. rep., Centre for Theoretical Neuroscience, Waterloo, ON.
- Gosmann, J., Voelker, A. R., & Eliasmith, C. (2017). A spiking independent accumulator model for winner-take-all computation. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*. London, UK: Cognitive Science Society.
URL <https://mindmodeling.org/cogsci2017/papers/0405/index.html>
- Gross, J., Schmitz, F., Schnitzler, I., Kessler, K., Shapiro, K., Hommel, B., & Schnitzler, A. (2004). Modulation of long-range neural synchrony reflects temporal limitations of visual attention in humans. *Proceedings of the national Academy of Sciences*, *101*(35), 13050–13055.
- Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia. i. a new functional anatomy. *Biological cybernetics*, *84*(6), 401–410.
- Head, T., MechCoder, G. L., Shcherbatyi, I., et al. (2018). scikit-optimize/scikit-optimize: v0. 5.2.
- Kim, C.-Y., & Blake, R. (2005). Psychophysical magic: rendering the visible ‘invisible’. *Trends in cognitive sciences*, *9*(8), 381–388.
- Mashour, G. A., Roelfsema, P., Changeux, J.-P., & Dehaene, S. (2020). Conscious processing and the global neuronal workspace hypothesis. *Neuron*, *105*(5), 776–798.
- Michel, M., Fleming, S. M., Lau, H., Lee, A. L., Martinez-Conde, S., Passingham, R. E., Peters, M. A., Rahnev, D., Sergent, C., & Liu, K. (2018). An informal internet survey on the current state of consciousness science. *Frontiers in psychology*, *9*, 2134.
- Miller, G. A. (1962). *Psychology: The science of mental life*. Harper & Row.
- Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology*, *8*(2), 228–246.

- Plate, T. A. (1995). Holographic reduced representations. *IEEE Transactions on Neural networks*, 6(3), 623–641.
- Redgrave, P., Prescott, T. J., & Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89(4), 1009–1023.
- Sackur, J., & Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*, 111(2), 187–211.
- Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature neuroscience*, 8(10), 1391–1400.
- Seth, A. K. (2018). Consciousness: The last 50 years (and the next). *Brain and neuroscience advances*, 2, 2398212818816019.
- Shanahan, M. (2008). A spiking neuron model of cortical broadcast and competition. *Consciousness and Cognition*, 17(1), 288–303.
- Shanahan, M. (2012). The brain’s connective core and its role in animal cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1603), 2704–2714.
- Shanahan, M., & Baars, B. (2005). Applying global workspace theory to the frame problem. *Cognition*, 98(2), 157–176.
- Stewart, T. C., Choo, X., & Eliasmith, C. (2010a). Symbolic reasoning in spiking neurons: A model of the cortex/basal ganglia/thalamus loop. In S. Ohlsson, & R. Catrambone (Eds.) *32nd Annual Meeting of the Cognitive Science Society*, (pp. 1100–1105). Portland, Oregon: Cognitive Science Society.
- Stewart, T. C., Choo, X., Eliasmith, C., et al. (2010b). Dynamic behaviour of a spiking model of action selection in the basal ganglia. In *Proceedings of the 10th international conference on cognitive modeling*, (pp. 235–40). Citeseer.
- Stewart, T. C., Penner-Wilger, M., Waring, R. J., & Anderson, M. L. (2017). A common neural component for finger gnosis and magnitude comparison. In *CogSci*.
- Stöckel, A., & Eliasmith, C. (2019). Passive nonlinear dendritic interactions as a general computational resource in functional spiking neural networks.
- Sutherland, N. S. (1989). *The international dictionary of psychology*. Crossroad Publishing Company.

- Thagard, P., & Stewart, T. C. (2014). Two theories of consciousness: Semantic pointer competition vs. information integration. *Consciousness and Cognition*, *30*, 73–90.
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. of Math*, *58*(345-363), 5.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, *108*(3), 550.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, Í., Feng, Y., Moore, E. W., Vand erPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., & Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272.
- Wyart, V., & Tallon-Baudry, C. (2009). How ongoing fluctuations in human visual cortex predict perceptual awareness: baseline shift versus decision bias. *Journal of Neuroscience*, *29*(27), 8715–8725.
- Zylberberg, A., Dehaene, S., Roelfsema, P. R., & Sigman, M. (2011). The human turing machine: a neural framework for mental programs. *Trends in cognitive sciences*, *15*(7), 293–300.
- Zylberberg, A., Slezak, D. F., Roelfsema, P. R., Dehaene, S., & Sigman, M. (2010). The brain's router: a cortical network model of serial processing in the primate brain. *PLoS Comput Biol*, *6*(4), e1000765.