A study on the spatial representation of numbers using context-dependent associative memories

by

Lic. Francisco López Guzmán

Tesis de Maestría
Maestría en Ciencias Cognitivas
Universidad de la República

Advisor
Dr. Andrés Pomi

Thesis comitee
Dr. Leonel Gómez
Dr. Alejandro Maiche
Dr. Juan Carlos Valle Lisboa

May 2021
Montevideo, Uruguay

Abstract

This thesis combines the theoretical framework of context-dependent associative memories with mathematical models of decision-making in an effort to better understand spatial-numerical associations in the mind. First, a computational model that allows the integration of multi-attribute evidence is proposed. Applying this model to experimental data from a color and motion random-dot kinetogram reveals a role for task-irrelevant information. The original model is subsequently extended to include spatial representations in order to reproduce the Simon effect. Finally, the model is applied to two experimental paradigms from the field of numerical cognition: the numerical distance effect and the SNARC effect. Overall, this work presents a comprehensive review of the relevant literature as well as a general theoretical framework for the processing of conflicting information in context-dependent decision tasks.

Resumen

Esta tesis combina el marco teórico de las memorias asociativas dependientes de contextos con modelos matemáticos de toma de decisiones en un esfuerzo por comprender las asociones número-espacio en la mente. En primer lugar, se propone un modelo computacional que permite la integración de evidencia multiatributo. La aplicación de este modelo a datos experimentales de un kinetograma con color y movimiento revela un rol de importancia para la información irrelevante en la tarea. El modelo original es luego extendido para incluir representaciones espaciales de forma de poder reproducir el efecto Simon. Finalmente, el modelo es aplicado a dos paradigmas experimentales del campo de la cognición numérica: el efecto de distancia numérica y el efecto SNARC. En general, este trabajo presenta una profunda revisión de la bibliografía relevante y un marco teórico para el procesamiento de información conflictiva en tareas de decisión dependientes de contexto.

Contents

List of Figures

List of Tables

List of Acronyms

AIP   Anterior intraparietal area

ANS   Approximate number system

BSB   Brain-state-in-a-box

DDM   Drift-diffusion model

DMC   Diffusion model for conflict tasks

EEG   Electroencephalography

fMRI   Functional magnetic resonance imaging

IPS   Intraparietal sulcus

LCA   Leaky, competing accumulator model

LIP   Lateral intraparietal area

LRP   Lateralized readiness potential

MDFT   Multi-alternative decision field theory

MIP   Medial intraparietal area

MLCA   Multi-attribute leaky, competing accumulator model

MNL   Mental number line

MT   Middle temporal area

OTS   Object tracking system

PFC   Prefrontal cortex

PPC   Posterior parietal cortex

RDK   Random-dot kinetogram

SNARC   Spatial-numerical association of response codes effect

SNL   Spatial number line

SRC   Stimulus-response compatibility

VIP   Ventral intraparietal area

Introduction

During the three decades following the first paper on context-dependent associative memories (Mizraji, 1989), several researchers have been inspired to implement this model in their work. Topics range from logical reasoning to language production and medical diagnosis. Each new individual brings their own preferences and incorporates novel techniques, thus modifying the model in their best interest. But all these works share a common core: the idea that context, seemingly of second-order importance or even irrelevant, actually plays a key role in the processing of information. I believe this idea can also be applied to what I consider one of the most fascinating observations within the field of numerical cognition: a shared representation for numbers and space. The present thesis uses the framework of context-dependent associative memories as a fundamental building block in an effort to understand spatial-numerical associations.

A priori, one might expect space and number to be independent magnitudes with correspondingly independent mental representations. In the late 1800s, Francis Galton reported that, when asked how they thought about numbers and arithmetic, some subjects visualized specific spatial arrangements. One of them expressed that "every number [...] is always thought of by me in its own definite place in the series, where it has, if I may say so, a home and an individuality" (Galton, 1880, p. 253). Another subject claimed: "Numerals are always pictured by me in a straight line from left to right. They are black, on a ground varying in illumination" (Galton, 1880, p.254). These peculiar shapes receive the name of number forms (Fig. I.1). Recent research reveals that the number forms reported by Galton are a form of synesthesia present in a small fraction of the population (Sagiv et al., 2006).

Nevertheless, the idea of a number line in the mind is nowadays accepted as universal. Restle (1970) performed a number comparison study to prove that an analogical mental number line is in fact a satisfactory representation of numbers, a result later supported by the tuning curves of number neurons in monkeys (e.g. Nieder and Miller, 2004). Stanislas Dehaene and two of his students found that, during a parity judgment task, left responses were faster for small numbers while right responses were faster for large numbers, even though numerical magnitude was task-irrelevant (Dehaene et al., 1993; Fig. I.2). They named this phenomenon the Spatial-Numerical Association of Response Codes (SNARC). The discovery of the SNARC effect sparked a sizable interest in the study of the mapping of numbers onto space. Many conflicting accounts have been proposed to explain this effect. Most of them assume a shared representation of numbers and space as if the mental number line had a spatial nature. However, recent experimental results appear to grapple with this possibility, such as the works of Wim Fias and collaborators (e.g. Fias and van Dijck, 2016) which highlight the importance of working memory and the task-specific associations between numerical magnitudes and lateralized responses.

Throughout the thesis, I develop a gradual account of the mental representation of numbers and space, with the objective of proposing a neurocomputational mechanism that can explain the decision process during a SNARC task. Experimentally, decision-making is researched both at the behavioral and neural levels, while numerous models intend to bridge the gap between the two. The present work builds on previous
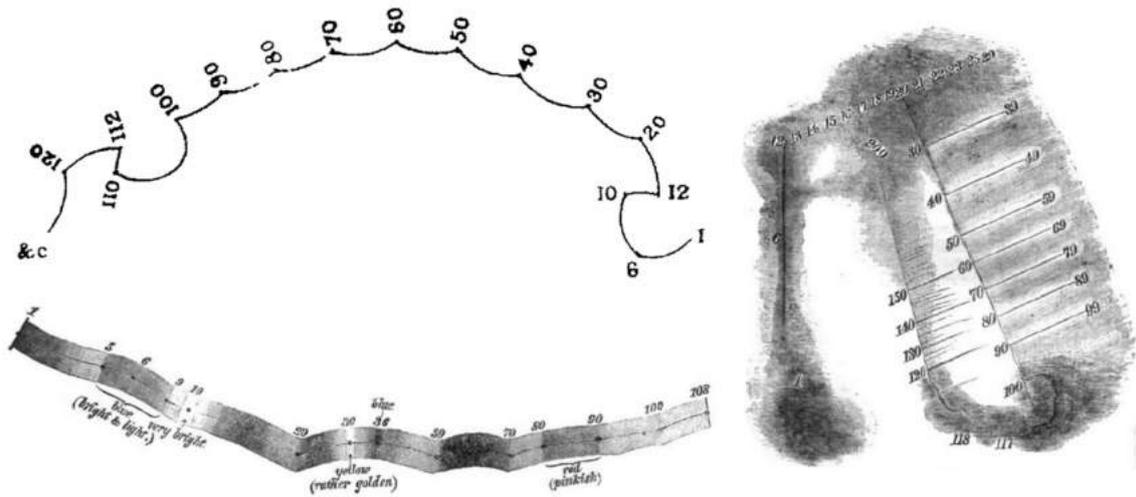
Figure I.1: Examples of number forms reported by three subjects (adapted from Galton, 1880). Apart from specific locations for each number, subjects also visualize these number forms as having different sizes, colors, and brightness.

mathematical models of decision-making that have been successfully used to reproduce laboratory data. The relevance of laboratory decision-making tasks to understand mental and neural representations is not obvious. Ideally, one would want to study cognition in ecological environments. This is, however, nearly impossible. By using controlled stimuli with limited attributes and limited responses, experimenters can pin down the causes for the effects they observe. The same goes for computational models. Such is the approach chosen for the present work.

According to Dehaene's triple-code model (Dehaene, 1992; Dehaene and Cohen, 1995), magnitude and parity are two independent attributes of a number. The first stage of this work, therefore, focuses on how humans and non-human animals make decisions based on stimuli composed of different attributes. In Chapter 1, I summarize some of the main experimental results from the field of perceptual decision-making and introduce a class of computational models, called sequential sampling models, that successfully reproduce said results. Rather than suggesting an entirely new one, my proposal looks to extend existing computational models into the representational domain. A detailed explanation of one of those models, the leaky, competing accumulator model (LCA; Usher and McClelland, 2001), is given. I then present an innovative vector-matrix formulation of the LCA and show how context-dependent associative memories allow the integration of multi-attribute evidence. In doing so, I am able to reproduce results that previous models had failed at. Chapter 1 ends with an application of the proposed model to the multi-attribute kinetogram study of Mante et al. (2013).

An important intermediate step in relating decision-making and the SNARC effect is the mapping of perceptions and actions into a common spatial domain. The second part of this thesis is dedicated to stimulus-response compatibility effects, and particularly to the Simon effect. Incompatibilities between a stimulus and its corresponding response imply slower and less accurate decisions. In the case of the Simon effect, these incompatibilities are due to differences in the hemispatial locations of the stimuli and the responses. Although the locations of the stimuli are task-irrelevant, neural and behavioral data indicate there is an automatic activation which leads to the impulse of executing the spatially-corresponding action. Chapter 2 introduces the idea of spatial representations for both stimuli and responses, which are added to the model presented in Chapter 1. A dual-route version of the multi-attribute model is able to reproduce the main experimental results of the Simon
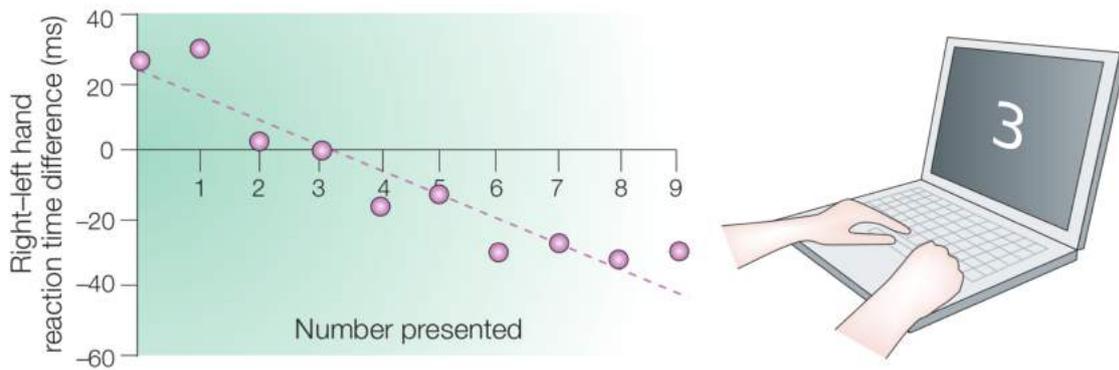
Figure I.2: Schematic overview of the SNARC effect (adapted from Hubbard et al., 2005). Participants are asked to indicate the parity of a number on a screen by pressing one of two buttons. Response times are faster for lower numbers with the left button and for larger numbers with the right button.

effect.

Finally, in Chapter 3, I focus on the relation between the representations of numbers and space. I begin with an overview of the behavioral, neural, and cultural evidence that supports the idea of a mental number line. Next, I present the experimental results that are used to argue in favor of a spatial mapping of numbers, namely the SNARC effect and its variants. Upon a revision of recent findings that challenge the hypothesis of a permanent shared representation for numbers and space, I suggest a theoretical account in which three different levels can accommodate all experimental findings: neural overlap, long-term associations, and short-term mappings. Chapter 3 finishes with applications of the model to numerical cognition. First, I reproduce the numerical distance effect to justify the use of sequential sampling models. I then perform numerical simulations with the multi-attribute version of the model to capture the SNARC and thus show that associations between numerical magnitudes and spatially defined responses are sufficient to explain the effect.

Computational models of decision-making, as most computational models in cognitive science, are usually critized for only being able to explain a specific phenomenon without any generalization capabilities. The goal for this thesis is not merely to develop a satisfactory account of the SNARC effect, but also to suggest a biologically plausible neurocomputational mechanism that relates the representations of numbers and other magnitudes with the intention of approximating the neural, computational and behavioral levels in numerical cognition and cognitive science in general. The extent to which this goal is achieved is discussed in the Conclusions.

Chapter 1

Perceptual decision-making with multiple attributes

In this Chapter, I put forward a model which allows the integration of multi-attribute evidence in perceptual decision-making. In Section 1.1, I present a summary of the most relevant experimental results from the field and the computational models which are used to reproduce laboratory data. In Section 1.2, I introduce a novel vector-matrix formulation of one of these models, the LCA of Usher and McClelland (2001), thus adding a distinction between stimuli representations and readouts based on a linear associative matrix. In Section 1.3, I extend the vector-matrix model to multiple attributes by using context-dependent associative memories. In Section 1.4, I test the capabilities of the model to reproduce the behavioral results of Mante et al. (2013) in a multi-attribute perceptual decision-making experimental paradigm, the color and motion kinetogram.

1.1. An overview of perceptual decision-making

Arguably one of the most prolific topics in cognitive science, decision-making is related to a plethora of behaviors. One could not claim to fully understand human or animal decision-making without satisfactorily explaining the role perception, response selection, motor execution, memory, biases, emotions, or any other of the numerous brain functions involved in the choices that we make constantly. Within the context of this thesis, the main interest is fast choices made in laboratory settings, where one of a limited number of actions must be executed in response to a controlled stimulus. These types of decisions are often grouped under the name of perceptual decision-making, defined as the cognitive function by which humans and non-human animals take a course of action based on the processing of sensory information. The difference between perceptual decision-making and other related fields of study, according to Hauser and Salinas (2013), is that the goal of the former is "to reveal the computational mechanisms whereby neural circuits encode, store, and analyze perceptual signals; combine them with other behaviorally relevant information; and use them to resolve conflicts between competing motor plans." Summerfield and Blangero (2017) also highlight the conversion from continuous sensory inputs to discrete categorical variables, a neurocomputational problem that involves the detection, discrimination and categorization of sensory information from the environment.
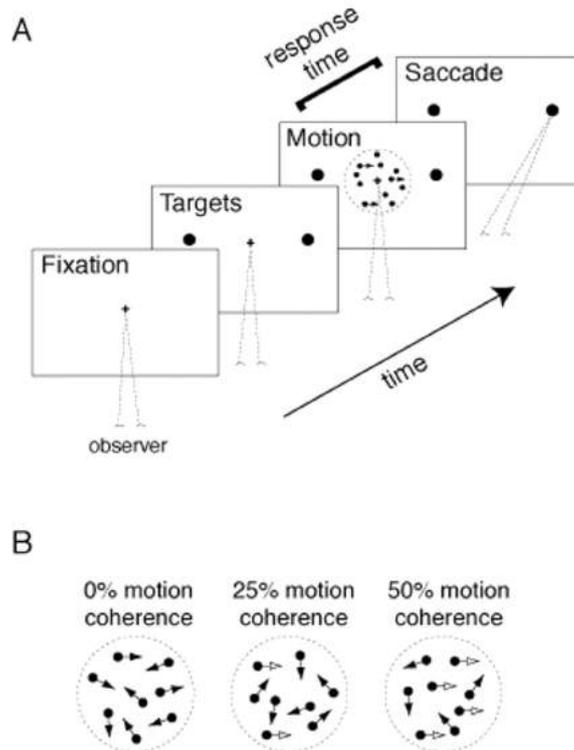
Figure 1.1: Example of a RDK task (Palmer et al., 2005). A: Timeline. Response time corresponds to the time between the presentation of the kinetogram and the saccade. B: Different motion coherences. Experimenters control the motion coherence as the percentage of dots that move in a fixed direction while the rest move randomly.

1.1.1.   Experimental results

Although the term "perceptual decision-making" was first used in the late 1980s (Newsome et al., 1989), one could argue that the field itself has been active at least since the 1940s. Those pioneering studies involved the psychophysical detection of sensory events, such as flashes of light in a dark environment, by humans (Hecht et al., 1942) or crabs (Hartline and McDonald, 1947). The seminal studies of the visual systems of frogs (Lettvin et al., 1959) and cats (Barlow et al., 1967; Hubel and Wiesel, 1962) revealed the neural origins of perceptual decisions from recordings of retinal ganglion cells as well as visual cortex. Parker and Newsome (1998) give a comprehensive review of the experimental developments made throughout the twentieth century.

Since the turn of the century, perceptual decision-making has been the subject of extensive psychophysical (Palmer et al., 2005; Kiani et al., 2014), physiological (Mazurek et al., 2003 Roitman and Shadlen, 2002; Gold and Shadlen, 2007; Rorie et al., 2010; Middlebrooks and Sommer, 2012), and neuroimaging experimentation (Heekeren et al., 2004; Wyart et al., 2012; Donner et al., 2009) in order to understand the computational mechanisms by which perceptual decisions are made. The field has mainly focused on relating the well-established psychophysics of perceptual decisions with the newly available physiological and neuroimaging data. To do so, a common experimental paradigm used in many studies is the random-dot kinetogram (RDK; Braddick, 1974; Baker Jr and Braddick, 1982). Subjects are presented with an animated cloud of apparently randomly moving dots, although a controlled fraction of these move coherently (Fig. 1.1). The task consists of indicating the net direction of motion, with humans usually having to press a button while monkeys perform saccades to a target. Other types of tasks related to perceptual decision-making include redundant multi-sensory stimuli detection (Diederich and Colonius, 2004),
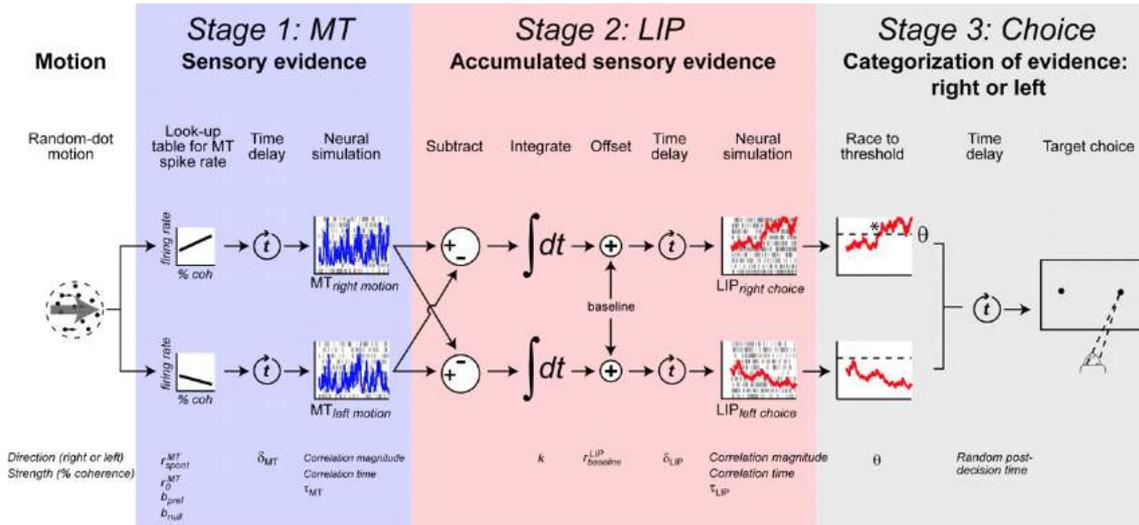
Figure 1.2: Three stages of perceptual decision-making (Mazurek et al., 2003). Physiological recordings indicate motion-selective neurons in MT feed evidence-accumulating neurons in LIP, which trigger a saccadic response once accumulation reaches a threshold.

speed perception (Sotiropoulos et al., 2014), and several perceptual effects such as the flash-lag effect (Maiche et al., 2007).

The work of Mazurek et al. (2003), who perform neural recordings on monkeys during a RDK task with saccadic responses, suggests perceptual decisions occur in three successive stages: first, the sensory evidence is encoded in the middle temporal area (MT) by motion-selective neurons; second, incoming evidence from MT is accumulated by neurons in the lateral intraparietal area (LIP) of the intraparietal sulcus (IPS) until there is enough evidence in favor of one response; finally, once the choice is made, the corresponding action commands are executed (Fig. 1.2). Other areas where evidence accumulation is detected in single cells include the medial intraparietal area (MIP; de Lafuente et al., 2015) and frontal eye field (Purcell et al., 2012) in monkeys, and the frontal orienting fields in rodents (Hanks et al., 2015). A similar accumulation process is observed in electroencephalography (EEG) recordings of human subjects during a RDK task (Kelly and O'Connell, 2013; Fig. 1.3). Both in single-cell and EEG data, the coherence of the stimulus determines how fast information is accumulated.

Palmer et al. (2005) study the effect of varying stimulus strength, such as the motion coherence of Fig. 1.1, on the speed and accuracy of decisions. Strong coherences are characteristically associated with more accurate responses, while weaker coherences imply higher error rates. The same applies to response times: the stronger the stimulus coherence, the faster the response time. The two are linked by a well-known phenomenon called the speed-accuracy tradeoff (Wickelgren, 1977) which experimenters can modulate by controlling instructions or response possibilities. Bogacz et al. (2010) attribute the neural basis of the speed-accuracy tradeoff to parietal and premotor areas during the integration stage.

Perceptual decision-making models must account for the relation between stimuli and behavior, particularly psychometric curves and the speed-accuracy tradeoff, and simultaneously be compatible with the accumulation observed with physiological and EEG recordings. However, a recent review by Najafi and Churchland (2018) claims the field is on the verge of major changes due to the development of new technologies and techniques, particularly concerning neural recordings. In coming years, monumental amounts of new data will either confirm the predictions of existing models or require the
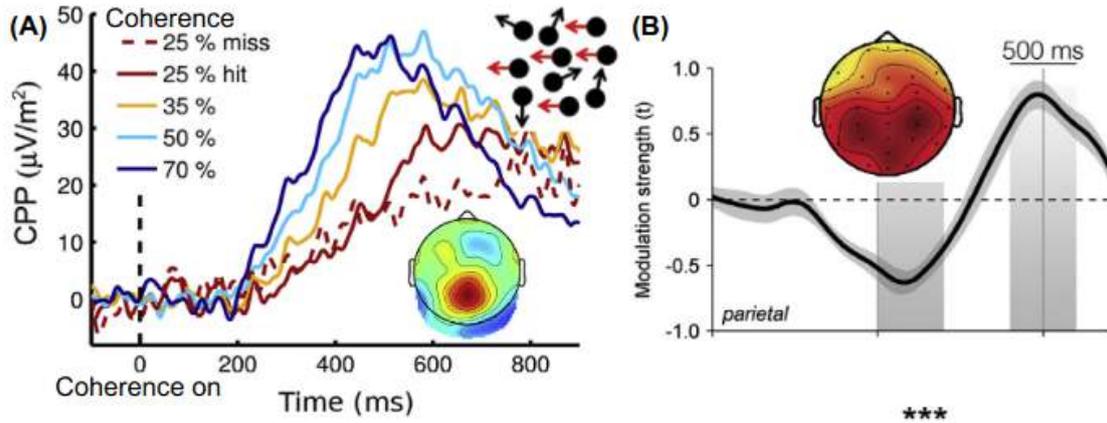
Figure 1.3: EEG recordings in humans reveal evidence accumulation in parietal cortex (Kelly and O'Connell, 2013). Motion coherence correlates with the speed of accumulation such that higher coherences lead to faster and more accurate responses.

inception of new ones.

1.1.2. Sequential sampling models

Different classes of models have been proposed to explain different aspects of decision-making. Bayesian models are useful to unveil prior biases for choices under uncertainty (e.g. Sotiropoulos et al., 2014); artificial neural networks can explain how humans learn to recognize patterns from experience (e.g. Krizhevsky et al., 2012). However, because the interest in this work is the computational mechanisms involved in the fast selection of an appropriate response upon perception of a stimulus, I decided to focus on a category of models called sequential sampling models (Busemeyer, 1985; Ratcliff and Smith, 2004). These are built on the idea that perceptual information is stochastically accumulated until one of the responses is selected.

The first sequential sampling models date back to the 1960s, when psychologists realized that the variability in response times and accuracy could be reproduced either by a continuous Wiener process or a discrete random walk (Stone, 1960; Laming, 1968). Vickers (1970) proposed a race model, where different alternatives are represented by independent variables that accumulate evidence until one "wins the race", i.e. reaches a threshold, and the corresponding response was selected. The most successful model from that era is the drift-diffusion model (DDM; Fig. 1.4), which can account for numerous two-alternative tasks and is still extensively used in psychophysical research (Ratcliff, 1978; Ratcliff and McKoon, 2008). The stochastic nature of the DDM is captured by the differential equation

$$dx = \mu \, dt/\tau + \xi\sqrt{dt/\tau} \tag{1.1}$$

where dx is the instantaneous evidence accumulated, dt is the time integration variable, μ is a drift coefficient, ξ is a Gaussian noise, and τ is a time scale for integration. The accumulation in (1.1) is the sum of two terms: the latter is a common diffusion of zero mean and a standard deviation which characteristically grows as the square root of time; the first term is an added drift which guides the accumulation variable x towards the positive or negative values, depending on the sign of the drift coefficient μ. Once x reaches either of the two boundaries (1.4), the accumulation process finishes and the corresponding response is selected. Although the DDM seems like a fairly simple model, its complexity and flexibility stems from the numerous other parameters that complete the integration

7
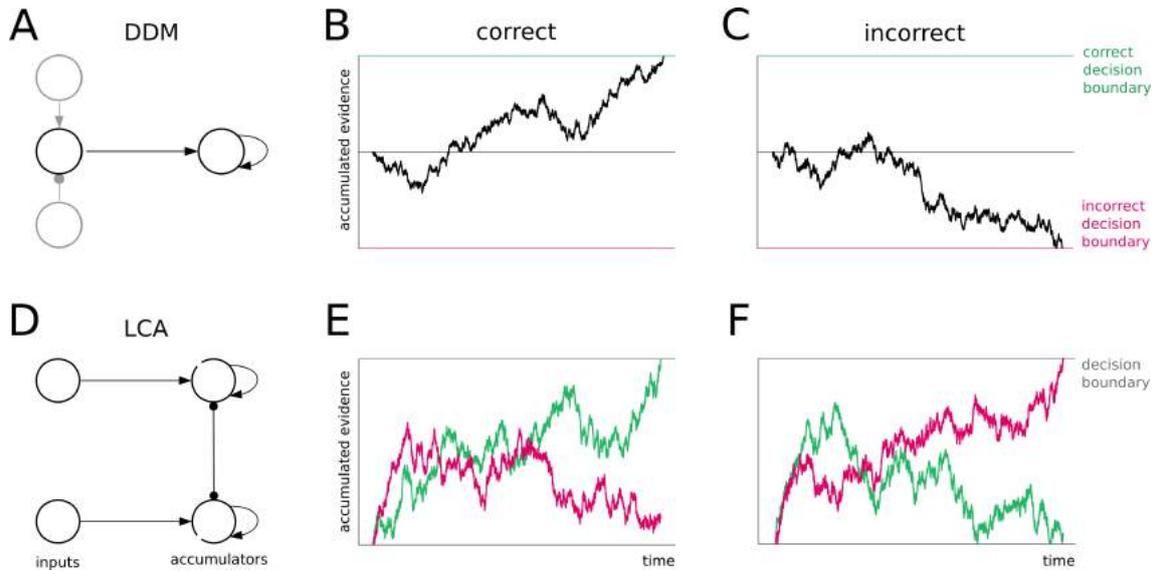
Figure 1.4: Overview of two perceptual decision-making models, the DDM (Ratcliff, 1978) and the LCA (Usher and McClelland, 2001). A: Architecture of the DDM with a single input, here represented as the difference of two implicit input nodes, and a single accumulator. B-C: Examples of the time evolution of accumulated evidence for correct and incorrect trials in the DDM. Accumulation finishes once the accumulated evidence reaches one of the decision boundaries. D: Architecture of the LCA model with two input nodes and two accumulator nodes with leakage and lateral inhibition. E-F: Examples of the time evolution of accumulated evidence for correct and incorrect trials in the LCA model. Accumulation finishes once one of the variables reaches the decision boundary.

process. These include the threshold values, which may be time-dependent, the starting point for the accumulation variable, and pre- and post-decision time parameters. These, along with the drift coefficient, are often sampled from distributions which need to be fitted to experimental data. Overall, the DDM is a useful computational tool for reproducing behavioral data, but it is unclear whether the model itself can be said to capture neural dynamics.

By the beginning of the 21st century, several physiological experiments had shown that the brain does in fact present an accumulation-like dynamic during the response selection stages of perceptual decision-making (see Section 1.1.1). However, researchers found different populations of neurons associated with each alternative, unlike the DDM. A response is selected only when the threshold is attained for the corresponding neurons' firing rates, regardless of the other populations (Shadlen and Newsome, 2001). Also, complex inhibitory dynamics are detected between neurons of different populations (Rorie et al., 2010).

These results led to a new generation of neurally-inspired computational models, which share a few key features but have their differences. Because of the existence of individual accumulators for each possible response, these models are often seen as more complex versions of the race model (Vickers, 1970; Vickers and Packer, 1982). Ditterich et al. (2003) and Mazurek et al. (2003) propose a feed-forward model, where each input node has an excitatory connection with its corresponding accumulator and inhibitory connections with the rest. Usher and McClelland (2001) use a mutual inhibition model, called the leaky, competing accumulator (LCA; Fig. 1.4), where input nodes only have excitatory connections but the accumulators have both leakage and mutual inhibitory connections. Wang (2002) includes an extra layer of pooled inhibitory nodes which deal
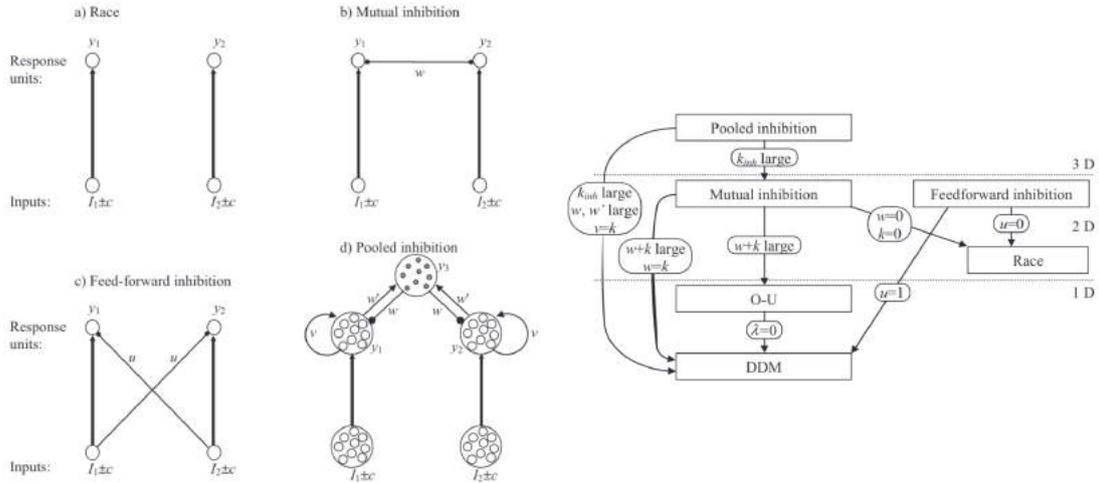
Figure 1.5: Main sequential sampling models of decision-making and how they relate to each other (Bogacz et al., 2006). Models shown include race (Vickers, 1970; Vickers and Packer, 1982), mutual inhibition (Usher and McClelland, 2001), feed-forward inhibition (Ditterich et al., 2003; Mazurek et al., 2003), pooled inhibition (Wang, 2002), and DDM (Ratcliff, 1978).

with the connections between the accumulator nodes. These models can be converted into each other and into the DDM (Bogacz et al., 2006; Fig. 1.5).

Although the aforementioned models have been successful at reproducing laboratory data, it is unclear how they could be extended to more ecological settings. Real-life decisions are seldom limited to a handful of possibilities; furthermore, they are usually taken based on a large quantity of multimodal sensory stimuli occurring simultaneously. Some efforts have been made to extend existing models to multi-alternative decisions (e.g., Usher and McClelland, 2001; Niwa and Ditterich, 2008; Churchland et al., 2008). Regarding multi-attribute inputs, most of the modeling effort has been focused on economic rather than perceptual decision-making (Usher and McClelland, 2004; Busemeyer and Townsend, 1993; Roe et al., 2001; Trueblood et al., 2014; Ronayne and Brown, 2017; Diederich, 1997, 2003; see Summerfield and Tsetsos (2012) for a comprehensive discussion on perceptual and economic decision-making). Diederich (2016) and Diederich and Oswald (2014) suggest multi-attribute stimuli are processed sequentially by means of an attention-switching process. The dynamics proposed are those of an Ornstein–Uhlenbeck process, which is equivalent to a single-variable LCA model (see Usher and McClelland, 2001), with a single attribute being used to select the response in each integration step.

An important yet often neglected component of decision-making is the distinction between the representation of a stimulus and its readout (Gold and Ding, 2013; Birman and Gardner, 2019). The representation is the neural activity associated with encoding the sensory input, such as high firing rates of direction-selective neurons in MT (Gold and Shadlen, 2007). This neural activity is characteristic of the stimulus and independent of the task. A readout, on the other hand, is the decoding of the stimulus representation by combining the activity of multiple neurons in order to select the appropriate response. Most perceptual decision-making models use representations and readouts interchangeably; it should be noted that readouts are task-specific and require some form of neural processing to associate each different stimulus to its corresponding response.

To this day, no single model is able to fully explain both behavioral and neural data (Churchland and Kiani, 2016; O'Connell et al., 2018).

1.1.3. The leaky, competing accumulator model

Out of all the existing computational models, I identified the LCA model (Usher and McClelland, 2001) as being the best candidate to use as the foundation for the novel mathematical developments introduced throughout this thesis. There are several reasons for this decision. First and foremost, as explained before, the LCA is one of the neurally-inspired computational models which incorporate the findings from electrophysiology. The present work is done with the aim of giving a biologically plausible account of perceptual decision-making, and it is therefore impossible to neglect the results from the field of systems neuroscience. Second, the LCA model is successful at reproducing experimental data (Usher and McClelland, 2001; Bogacz et al., 2007; Tsetsos et al., 2012). Granted, it is not as widespread in the mathematical psychology community as the DDM, likely due to its higher complexity, but when preferred over other alternatives it achieves the expected results. Finally, the mathematics of the LCA are elegant and readily convertible to a vector-matrix formulation, which is the first step of my proposed extension of the model. Rather than coming up with a completely new computational mechanism for the accumulation of evidence, I decided to work with this tried-and-tested one.

The dynamics of the LCA model for a task involving two alternatives, as defined by Usher and McClelland (2001), are given by the pair of coupled differential equations

$$
\begin{aligned}
dx_1 &= \left[\mu_1 - \alpha x_1 - \beta x_2\right]\frac{dt}{\tau} + \xi_1\sqrt{\frac{dt}{\tau}} \\
dx_2 &= \left[\mu_2 - \alpha x_2 - \beta x_1\right]\frac{dt}{\tau} + \xi_2\sqrt{\frac{dt}{\tau}}
\end{aligned}
\tag{1.2}
$$

where the variables x_1 and x_2 measure the accumulated evidence for each of the two alternatives due to constant inputs μ_1 and μ_2 with Gaussian noises ξ_1 and ξ_2 of zero mean and variance σ^2; α and β measure leakage and lateral inhibition, respectively, and τ is a characteristic time scale for integration. Additionally, it is useful to work under the assumption of non-linearity, whereby accumulators do not take negative values, i.e. $x_i(t) = \max(x_i(t), 0)$. The non-linear LCA is both biologically and computationally more relevant than the linear version, as detailed by Bogacz et al. (2007).

In the simplest implementation of the LCA model, the accumulators are initially set to 0 and evolve according to (1.2). In a task under standard conditions, the selected response corresponds to the accumulator that first reaches a threshold boundary (Fig. 1.4). Alternatively, time-controlled tasks force subjects to make a decision when a certain time has elapsed, so that the highest accumulator is the one chosen. Other variations, such as random starting-points for the accumulators and trial-to-trial variability in input values or threshold boundaries, although interesting, are not discussed in the present work.

To better understand the dynamics of the LCA model, Bogacz et al. (2007) suggest a transformation from the original variables x_1 and x_2 into $y_1 = (x_1 - x_2)/\sqrt{2}$ and $y_2 = (x_1 + x_2)/\sqrt{2}$, such that (1.2) becomes

$$
\begin{aligned}
dy_1 &= \left(\frac{\mu_1 - \mu_2}{\sqrt{2}} + (\alpha - \beta)y_1\right)\frac{dt}{\tau} + \left(\frac{\xi_1}{\sqrt{2}} - \frac{\xi_2}{\sqrt{2}}\right)\sqrt{\frac{dt}{\tau}} \\
dy_2 &= \left(\frac{\mu_1 + \mu_2}{\sqrt{2}} + (-\alpha - \beta)y_2\right)\frac{dt}{\tau} + \left(\frac{\xi_1}{\sqrt{2}} + \frac{\xi_2}{\sqrt{2}}\right)\sqrt{\frac{dt}{\tau}}
\end{aligned}
\tag{1.3}
$$

Unlike (1.2), (1.3) is now a pair of uncoupled differential equations, so the dynamic evolutions of y_1 and y_2 are independent. In fact, each of these is an Ornstein–Uhlenbeck equation. Because the inputs μ_1 and μ_2 are positive, y_2 has faster dynamics than y_1. Numerical simulations performed by Bogacz et al. (2007) show an attractor line in the state space of the accumulators (Fig. 1.6). They call it the decision line, because a decision is made when the solution reaches one of its ends. Solutions are quickly drawn to
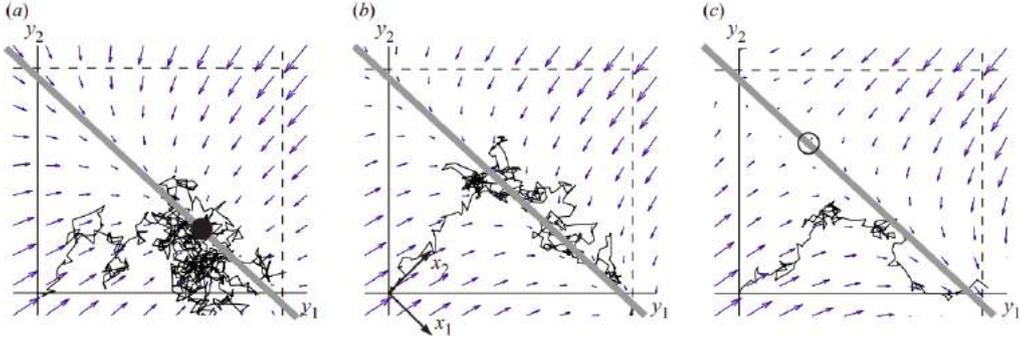
Figure 1.6: Dynamics of LCA in state space (Bogacz et al., 2007). (a): leakage greater than lateral inhibition leads to an attractor point. (b): leakage equal to lateral inhibition leads to a balanced model with no equilibrium points. (c): leakage smaller than lateral inhibition leads to a repeller point.

the decision line, and the evolution is then restricted to it, other than fluctuations due to random noise. The distance from that line to the origin is such that, neglecting the noise term, $dy_2(y_2^*) = 0$; that is $y_2^* = \sqrt{2}(\alpha + \beta)/(\mu_1 + \mu_2)$.

The dynamics parallel to the decision line are defined by y_1, and depend on the ratio of leakage to lateral inhibition. If $\alpha > \beta$, i.e. if leakage is greater than lateral inhibition, then there is an attractor point on the line to which solutions converge. If $\alpha < \beta$, i.e. if lateral inhibition is greater than leakage, then there is a repeller point from which solutions diverge. Usher and McClelland (2001) associate these two conditions as being responsible for a recency effect and a primacy effect, respectively. Tsetsos et al. (2012) further study these two effects with time-varying evidence inputs μ_1 and μ_2. They do so by switching the dominant input midway through the accumulation process, e.g. $\mu_1 > \mu_2$ for $0 < t < T/2$ and $\mu_1 < \mu_2$ for $T/2 < t < T$. When leakage is higher, old accumulated evidence is "forgotten" quickly, so that newer evidence plays a higher role in the decision-making process. The first half of the integration process has little effect on what happens during the second half, and the system is said to have no memory. This is referred to as a recency effect (Busemeyer and Townsend, 1993). Alternatively, when lateral inhibition is higher, evidence accumulated during the first half inhibits the accumulation of new evidence during the second half. This is called a primacy effect, such that the first information that arrives determines the evolution of the system. In state space, Bogacz et al. (2007) attribute recency and primacy to the attractor and repeller fixed points.

If $\alpha = \beta$, i.e. if leakage and lateral inhibition are equal, there are no equilibrium points and the model is said to be balanced. In this case, (1.3) and (1.1) become equivalent. The DDM can be seen as a special case of the LCA, where $\alpha = \beta$ and the dynamics are restricted to the decision line. Bogacz et al. (2007) show that the balanced model is close to optimality, in the sense that the model obtains the fastest response times for a fixed error rate. The DDM and the non-linear, balanced LCA are computationally equivalent, although the DDM is usually preferred for having fewer parameters and a single decision variable. As explained before, I decided to base my model on the LCA.

1.2. On the representation of stimuli

Consider two stimuli, a and b, such that when stimulus a is shown, the expected response corresponds to accumulator x_1 and when stimulus b is shown it corresponds to x_2. Subjects may occasionally make mistakes, captured by the inclusion of step-by-step noise. Assuming there are no further difficulties involved in the task, the following conditions should be met: $\mu_1^{(a)} > \mu_2^{(a)}$ and $\mu_1^{(b)} < \mu_2^{(b)}$. Usher and McClelland (2001) suggest an extra constraint: that the total information distributed over the accumulators should be kept constant for all stimuli. This is done by imposing $\mu_1 + \mu_2 = 1$, with 1 being an *a priori* arbitrary value related to the other free parameters. It is therefore necessary that stimuli a and b satisfy $0.5 < \mu_1^{(a)}, \mu_2^{(b)} \leq 1$, with evidence values close to 1 leading to shorter response times and fewer errors.

Given that the values of μ_1 and $\mu_2 = 1 - \mu_1$ are adjusted so that the model fits the corresponding experimental data, one could infer that these parameters indicate certain characteristics of the stimuli. However, the LCA model is indifferent to the representation of stimuli in possible previous layers and, due to its phenomenological nature, does not make assumptions about the representation of stimuli in the brain either. Next, I show how a vector representation of the neural variables and the use of linear associative memories allow the explicit inclusion of stimuli within the LCA model. Then, I derive a series of restrictions on both the stimuli and the associative memory.

1.2.1. A vector-matrix formulation of LCA

Incorporating vector variables into solid computational models is a must, if we want to understand how real brains compute. The vector representation relies on the fact that any kind of information traveling inside the brain is coded by the activity of large groups of neurons (Anderson, 1995; Friston, 2011). Large-dimensional vectors are, therefore, natural mathematical variables to represent neural activity.

The transformation of the LCA model into a vector-matrix model is straightforward. Let $\boldsymbol{x} = [x_1, x_2]^T$, $\boldsymbol{dx} = [dx_1, dx_2]^T$, $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$, $\boldsymbol{\xi} = [\xi_1, \xi_2]^T$ be column vectors. It can be readily seen that (1.2) can be rewritten as a vector-matrix equation:

$$\boldsymbol{dx} = [\boldsymbol{\mu} - \boldsymbol{Dx}] \frac{dt}{\tau} + \boldsymbol{\xi} \sqrt{\frac{dt}{\tau}} \qquad (1.4)$$

where $\boldsymbol{D}$ is a bisymmetric square matrix with positive entries, which I call the decay matrix, given by

$$\boldsymbol{D} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \qquad (1.5)$$

The balanced LCA, in which lateral inhibition and leakage are equal, achieves approximately optimal performance (Bogacz et al., 2007). In that case, $\alpha = \beta$, and therefore $\boldsymbol{D} = \alpha \boldsymbol{A}$, where $\boldsymbol{A}$ is an all-ones matrix. For the time being, I do not assume α and β to be equal.

(1.4) holds for the general LCA model with m multiple accumulators, where vectors $\boldsymbol{x}$, $\boldsymbol{dx}$, $\boldsymbol{\mu}$, and $\boldsymbol{\xi}$ are m-dimensional and the decay matrix $\boldsymbol{D}$ has dimensions $m \times m$ with values α in the main diagonal and β elsewhere. The term $[\boldsymbol{\mu} - \boldsymbol{Dx}]$ indicates the noise-free dynamics. Initially, with $\boldsymbol{x} = 0$, accumulation is proportional to input $\boldsymbol{\mu}$. As $\boldsymbol{x}$ changes so does $\boldsymbol{dx}$. Since $\boldsymbol{D}$ is bisymmetric, if $\alpha \neq \beta$ then its determinant is non-zero and there is a unique equilibrium $\boldsymbol{x_{eq}}$ such that $\boldsymbol{\mu} = \boldsymbol{Dx_{eq}}$ which, if $\alpha > \beta$, is an attractor point; otherwise it is a repeller. Alternatively, if $\alpha = \beta$, the determinant is zero and there are either infinite equilibrium points if evidence for all accumulators is the same, or there are

no equilibrium points. These are the same results obtained by Bogacz et al. (2007) with a transformation of variables, as detailed in Section 1.1.3.

1.2.2. Linear readouts

In their original paper, Usher and McClelland (2001) mention that their model is inspired by a two-layer architecture, with input nodes in the first layer and accumulator nodes in the second one. However, for simplicity, they neglect the first layer by using scalar inputs μ_1 and μ_2 instead. A distinction between a first layer of general-purpose representation nodes and their task-specific readout and accumulation in a second layer (Gold and Ding, 2013) is necessary to extend any computational model to the representational domain (Marr, 1982). I make explicit input-output associations between the stimuli and their readouts with a linear associative memory (Anderson, 1972; Cooper, 1973; Kohonen, 1972, 1977).

Associative memories are one of the earliest artificial neural network models inspired by the content-addressable nature of information storage in the brain. These systems provide a distributed mapping from inputs onto outputs. A matrix $\boldsymbol{M}$ is called an associative memory because of its ability to store and retrieve information by means of associations (Anderson, 1972; Kohonen, 1977). The associated pairs of neural activity patterns are superposed on the same synaptic coefficients, represented by the entries of the matrix. Consider an example of a simple two-alternative experiment with two possible stimuli, a picture of a dog or a picture of a cat, and two corresponding responses, a left or a right button press. I define the vector codes $\boldsymbol{s^{(1)}}$ and $\boldsymbol{s^{(2)}}$ for the stimuli, and $\boldsymbol{r^{(1)}}$ and $\boldsymbol{r^{(2)}}$ for the responses. A simple associative matrix memory storing the correct pairs of stimulus-response vectors can be constructed with the outer product rule:

$$\boldsymbol{M} = \boldsymbol{r^{(1)}s^{(1)}}^T + \boldsymbol{r^{(2)}s^{(2)}}^T \tag{1.6}$$

If the stimuli are orthogonal, i.e. $\langle \boldsymbol{s^{(1)}}, \boldsymbol{s^{(2)}} \rangle = 0$, then the product of the matrix with one of the stimuli is proportional to the associated response vector:

$$\begin{aligned} \boldsymbol{Ms^{(1)}} &= \boldsymbol{r^{(1)}} \langle \boldsymbol{s^{(1)}}, \boldsymbol{s^{(1)}} \rangle + \boldsymbol{r^{(2)}} \langle \boldsymbol{s^{(2)}}, \boldsymbol{s^{(1)}} \rangle \propto \boldsymbol{r^{(1)}} \\ \boldsymbol{Ms^{(2)}} &= \boldsymbol{r^{(1)}} \langle \boldsymbol{s^{(1)}}, \boldsymbol{s^{(2)}} \rangle + \boldsymbol{r^{(2)}} \langle \boldsymbol{s^{(2)}}, \boldsymbol{s^{(2)}} \rangle \propto \boldsymbol{r^{(2)}} \end{aligned} \tag{1.7}$$

In this sense, $\boldsymbol{M}$ acts as a memory that recovers the correct response for each stimulus. Stimuli, which are generally not strictly orthogonal, may generate some degree of interference in this simple model. For linearly independent inputs, a perfect associative recall is possible (Kohonen, 1977). In general, the storage capacity of an associative memory is limited and is most efficient for sparsely coded input and output vectors (Liang and Bose, 1996). (1.6) is a fairly simple example of Hebbian learning (Hebb, 1949); other learning mechanisms (e.g. the delta rule) are also often used in computer science. However, (1.6) is a good enough approximation if one considers the stored stimuli to be abstract representations extracted from all the stimuli associated with each response. Returning to the aforementioned experiment, subjects may learn with different pictures of dogs and extract an abstract idea, so that when seeing a new picture of a dog, regardless of its breed, the associated response is to press the left button.

By incorporating an associative memory into the LCA equation (1.2), the dynamics of an accumulator vector $\boldsymbol{x}$ induced by the presence of a stimulus $\boldsymbol{s}$ are given by

$$d\boldsymbol{x} = [\boldsymbol{Ms} - \boldsymbol{Dx}] \frac{dt}{\tau} + \boldsymbol{\xi} \sqrt{\frac{dt}{\tau}} \tag{1.8}$$

where matrix $\boldsymbol{M}$ is the stimuli-readouts associator. It should be noted that this is simply a change of notation motivated by our interest in studying the representation of stimuli: the original LCA model is recovered by making $\boldsymbol{\mu} = \boldsymbol{Ms}$.

The representation $\boldsymbol{s}$ describes the neural activity associated with the encoding of the sensory input. This representation is decoded into a decision variable, called a readout $\boldsymbol{\mu}$, which is accumulated to select the correct response (Gold and Ding, 2013; Gold and Shadlen, 2007). The linear association between representations and readouts can be thought of as a first-order approximation; this will be seen to be enough for single-attribute decisions.

While $\boldsymbol{D}$ remains a square matrix, $\boldsymbol{M}_{(m \times m')}$ can be rectangular, since it stores a mapping from an m'-dimensional stimuli space onto the m-dimensional accumulators space. For the moment, I assume that $\boldsymbol{M}$ is given and fixed during an experiment. In the case of a two-alternative choice, one can immediately notice that $\boldsymbol{s}$ cannot be a scalar, for otherwise it would not be possible for two different stimuli to produce greater evidence in different accumulators. Stimuli should be at least two-dimensional vectors. This can be extended to any m-dimensional choice, where stimuli should be represented by m'-dimensional vectors, with $m' \geq m$.

The same way as each accumulator is considered to capture average firing rates of entire populations of parietal or prefrontal neurons, I suggest that stimuli vectors represent distributed firing rates of large groups of neurons from modality- and attribute-specific brain regions (Mazurek et al., 2003). It is reasonable then to assume that m' can be as large as necessary although, as seen next, a dimension $m' = m$ is sufficient for adequate discrimination.

Consider a two-alternative choice experiment, with m'-dimensional stimuli. The matrix $\boldsymbol{M}$ can be written as a $(2 \times m')$-partitioned matrix

$$\boldsymbol{M} = \begin{pmatrix} m_{11}\ m_{12}\ ...\ m_{1m'} \\ m_{21}\ m_{22}\ ...\ m_{2m'} \end{pmatrix} = \begin{pmatrix} \boldsymbol{m_1}^T \\ \boldsymbol{m_2}^T \end{pmatrix} \tag{1.9}$$

where $\boldsymbol{m_1}$ and $\boldsymbol{m_2}$ are m'-dimensional vectors whose components are the entries of each row of the matrix. Given a stimulus $\boldsymbol{s}$, the evidence arriving at each accumulator is $\mu_1 = \langle \boldsymbol{m_1}, \boldsymbol{s} \rangle$ and $\mu_2 = \langle \boldsymbol{m_2}, \boldsymbol{s} \rangle$. But $\boldsymbol{m_1}$ and $\boldsymbol{m_2}$ must be linearly independent, otherwise there would be no difference between the accumulators. Then, there is a unique two-dimensional subspace (i.e. a plane) that contains both vectors. This means that although the stimuli $\boldsymbol{s}$ need not be contained in that subspace, the model is exclusively concerned with their projections on it. Accordingly, for a two-alternative choice experiment, it is possible to define a new frame of reference for the stimuli space restricted to the subspace formed by $\boldsymbol{m_1}$ and $\boldsymbol{m_2}$. Generalizing this result, for m-alternative choices stimuli can be represented with m-dimensional vectors. Hence, for mathematical simplicity and without loss of generality, from here on I assume that $m' = m$ and therefore $\boldsymbol{M}$ is an $(m \times m)$ square matrix.

In order to better understand how this associative memory relates to known behavior, I now show its equivalent within the DDM. Recall from section 1.1.3 that a DDM model can be recovered from a 2-alternative LCA by setting $\alpha = \beta$; I also define a new decision variable $y = x_1 - x_2$ as the difference between the accumulators. Under these conditions, and naming $\boldsymbol{\Delta m} = \boldsymbol{m_1} - \boldsymbol{m_2}$, (1.8) becomes

$$\begin{aligned} dy &= dx_1 - dx_2 \\ &= \left(\langle \boldsymbol{m_1}, \boldsymbol{s} \rangle - \langle \boldsymbol{m_2}, \boldsymbol{s} \rangle \right) dt/\tau + \xi \sqrt{dt/\tau} \\ &= \langle \boldsymbol{\Delta m}, \boldsymbol{s} \rangle dt/\tau + \xi \sqrt{dt/\tau} \\ &= \|\boldsymbol{\Delta m}\| \cos(\theta)\ dt/\tau + \xi \sqrt{dt/\tau} \end{aligned} \tag{1.10}$$

where I also define θ as the angle between $\boldsymbol{\Delta m}$ and $\boldsymbol{s}$. Importantly, I assume all stimuli are normalized. (1.10) is a vector-matrix reformulation of the DDM equation (Ratcliff, 1978) with a drift coefficient μ equal to the difference between the readouts of the stimulus for

each alternative. When fitting the DDM to experimental data, one obtains a series of drift coefficients for each stimulus category or coherence. By arbitrarily setting $\|\boldsymbol{\Delta m}\| = \mu_{max}$ as the largest of these drift coefficients, the vector-matrix formulation of the DDM is shown to have the same number of parameters as the original DDM, with the new parameter θ and the original parameter μ related by $\cos(\theta) = \mu/\mu_{max}$.

However, unlike the original DDM, there is a natural interpretation of the drift coefficient: it is the degree to which the representation of the stimulus being perceived during an experimental trial corresponds to the representations stored in the associative memory. From (1.10), a high, positive drift means the stimulus is aligned with $\boldsymbol{\Delta m}$; a high, negative drift means the stimulus is antialigned with $\boldsymbol{\Delta m}$; and a low drift means the two vectors are nearly orthogonal. Transforming back from the decision variable y into the accumulators x_1 and x_2, these conditions correspond to a stimulus aligned with either $\boldsymbol{m_1}$, $\boldsymbol{m_2}$, or an ambiguous stimulus. Interestingly, within LCA, it would be possible for a stimulus to be similar to both $\boldsymbol{m_1}$ and $\boldsymbol{m_2}$, provided these two were also similar between them. This is a completely different scenario from a stimulus that is different from both stored vectors, even though they would be both captured by the same parameters in the DDM. This highlights the interest in using representational-level models that go beyond the basic computational approach.

1.3.   Multi-attribute evidence accumulation

When decision-making is based on evidence arising from multiple attributes, the observed behaviors are complex. Contextual changes can originate qualitatively different choices in the presence of the same stimulus, a phenomenon called "preference reversal" (Tsetsos et al., 2010). The modeling of this type of behavior requires models capable of capturing non-linear dynamics.

To achieve this goal, Usher and McClelland (2004) add a multi-layer architecture of nodes that preprocess inputs into a conventional LCA model. Their proposal involves excitatory and inhibitory projections from an input-item layer into a difference-input layer that in turn transmits information into the multiple accumulators. This model achieves satisfactory results, comparable to those of other mathematical models such as Multi-alternative Decision Field Theory (MDFT; Diederich, 1997; Roe et al., 2001).

I am interested in extending the vector-matrix formulation of the LCA model to account for multi-attribute inputs while simultaneously maintaining the neurobiological distributed representation. But classical linear associative memories share with single layer perceptrons the inability to capture nonlinear behaviors, a troublesome restriction linked to their impossibility to compute the XOR (Minsky and Papert, 1969) and to perform context-dependent associations, necessary to solve "the bifurcation problem" (Anderson, 2003).

1.3.1.   Context-dependent associative memories

An alternative to neural models with non-linear units and hidden layers, able to solve the context problem with additive key and context vectors, arises from matrix neural models with key and context vectors composed by a Kronecker product (Mizraji, 1989; Smolensky, 1990). These non-linear associative memories, called context-dependent associative memories (Mizraji, 1989) allow the development of theoretical approaches for complex cognitive activities using the power of matrix algebra (beim Graben and Potthast, 2009). These vector states models capture complex cognitive functions such as the selective extraction of features from complex patterns (Pomi and Mizraji, 1999), adaptive searching (Mizraji et al., 2009) or logical reasoning (Mizraji and Lin, 1997, 2002, 2011). The thematic packaging of information, an essential feature of context-dependent associative memories, shows a kinship between these models and the vector space models used to extract information from databases as it is the case with latent semantics (Mizraji et al., 2009) and with search engines (Mizraji, 2008). Other applications of this formalism include the organization of thematically ordered sequences in the production of language (Valle-Lisboa et al., 2014; Mizraji, 2008), the exploration of the cognitive processes involved in medical diagnosis (Pomi and Olivera, 2006; Pomi, 2017) and the possibility of spatial organizations of memories in semantic topographies (Pomi et al., 2018). With their vector-matrix algebraic nature, these associative memory models with tensor contexts provide a unified framework for developing neural based cognitive theories.

Non-linearity is achieved by taking the tensor Kronecker product (see Graham, 1981) of two or more vectors representing stimuli and potentially contexts. The Kronecker product, $\otimes$, is a particular case of the tensor product of two matrices of arbitrary dimensions, where $\boldsymbol{A} \otimes \boldsymbol{B} = [a_{ij}\boldsymbol{B}]$. The following are some of its basic properties:

$$
\begin{aligned}
&\text{i.} \quad \lambda(\boldsymbol{A} \otimes \boldsymbol{B}) = \boldsymbol{A} \otimes (\lambda \boldsymbol{B}) \\
&\text{ii.} \quad \boldsymbol{A} \otimes \boldsymbol{B} + \boldsymbol{A} \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} + \boldsymbol{C}) \\
&\text{iii.} \quad (\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = (\boldsymbol{A}\boldsymbol{C}) \otimes (\boldsymbol{B}\boldsymbol{D}) \\
&\text{iv.} \quad (\boldsymbol{A} \otimes \boldsymbol{B})^T = (\boldsymbol{A}^T \otimes \boldsymbol{B}^T)
\end{aligned}
\tag{1.11}
$$

Of importance for context-dependent memory models, if $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$, and $\boldsymbol{d}$ are k-dimensional column vectors, using properties (1.11.iii) and (1.11.iv), the following equations hold:

$$(\boldsymbol{a} \otimes \boldsymbol{b})^T (\boldsymbol{c} \otimes \boldsymbol{d}) = (\boldsymbol{a}^T \otimes \boldsymbol{c})(\boldsymbol{b}^T \otimes \boldsymbol{d}) = \langle \boldsymbol{a}, \boldsymbol{c} \rangle \langle \boldsymbol{b}, \boldsymbol{d} \rangle \tag{1.12}$$

Consider initially an m-alternative task with stimuli composed of n independent attributes, represented by vectors $\boldsymbol{s_1}, \boldsymbol{s_2}, ..., \boldsymbol{s_n}$ of dimensions $d_1, d_2, ..., d_n$. By the reasoning presented in Section 1.2.2, $d_1 = d_2 = ... = d_n = m$. The stimulus is thus coded by a single vector of dimension m^n formed by the multiple Kronecker product

$$\boldsymbol{s} = \boldsymbol{s_1} \otimes \boldsymbol{s_2} \otimes ... \otimes \boldsymbol{s_n} = \bigotimes_{i=1}^{n} \boldsymbol{s_i} \tag{1.13}$$

where each $\boldsymbol{s_i}$ represents a different attribute. Substituting into (1.8) gives a multi-attribute extension of the LCA model, which I henceforth call the multi-attribute, leaky, competing accumulator (MLCA) model:

$$\boldsymbol{dx} = \left[\boldsymbol{M} \left(\bigotimes_{i=1}^{n} \boldsymbol{s_i} \right) - \boldsymbol{Dx} \right] \frac{dt}{\tau} + \boldsymbol{\xi} \sqrt{\frac{dt}{\tau}} \tag{1.14}$$

As before, the decay matrix is a bisymmetric square matrix with leakage terms in the main diagonal and lateral inhibition terms elsewhere

$$\boldsymbol{D} = \begin{pmatrix} \alpha & \beta & ... & \beta \\ \beta & \alpha & ... & \beta \\ ... & ... & ... & ... \\ \beta & \beta & ... & \alpha \end{pmatrix} \tag{1.15}$$

The associative memory stores associations between the attributes that form each stimulus and the corresponding responses. Because of the Kronecker product, these associations are no longer linear and $\boldsymbol{M}$ is called a context-dependent associative memory (Mizraji, 1989) given by

$$\boldsymbol{M} = \sum_{j \in \mathcal{J}} \eta_j \boldsymbol{r^{(j)}} \left(\bigotimes_{i=1}^{n} \boldsymbol{s_i^{(j)}} \right)^T \tag{1.16}$$

where the sum ranges over all combinations from a set $\mathcal{J}$ of attributes and their associated responses $\boldsymbol{r^{(j)}}$, with scaling rates η_j. In a first approximation, I consider that this set is the one presented to the subjects during the training phase of an experiment, with no interference from previously stored stimulus-response pairings. This assumption need not be valid in a general case. Several perceptual and cognitive effects, such as the Simon effect (Simon and Small, 1969; see Chapter 2) and the SNARC effect (Dehaene et al., 1993; see Chapter 3), provide evidence of task-irrelevant associations between stimuli and responses. I also assume that the response vectors form an orthonormal standard basis in the accumulator space and there exists a set $\mathcal{J}^*$ that indexes each possible response once:

$$\forall (j, j') \in \mathcal{J}^*, \ \langle \boldsymbol{r^{(j)}}, \boldsymbol{r^{(j')}} \rangle = \delta_{jj'} \tag{1.17}$$

The context-dependent associative memory is therefore defined over the set $\mathcal{J}^*$, with the single stimulus stored for each response being an abstract generalization from the originally shown stimuli corresponding to that response. Without loss of generality, I consider $\mathcal{J}^*$ to be an ordered set of natural numbers from 1 to m such that each $\boldsymbol{r^{(j)}}$ corresponds to accumulator x_j.

During an experimental trial, a multi-attribute stimulus a is presented to the subject. The corresponding stimulus vector $\boldsymbol{s}^{(a)} = \bigotimes \boldsymbol{s}_k^{(a)}$ encodes the neural activation produced by that stimulus across all attributes. The input to the accumulators is the product of the associative memory $\boldsymbol{M}$ and the stimulus vector $\boldsymbol{s}^{(a)}$, given by

$$\boldsymbol{M}\boldsymbol{s}^{(a)} = \sum_{j=1}^{m} \eta_j \boldsymbol{r}^{(j)} \left(\bigotimes_{i=1}^{n} \boldsymbol{s}_i^{(j)}\right)^T \left(\bigotimes_{k=1}^{n} \boldsymbol{s}_k^{(a)}\right) \tag{1.18}$$

From the properties of the Kronecker product enumerated before, it can be seen that the product of the vectors is the scalar product of their pairwise inner products:

$$\boldsymbol{M}\boldsymbol{s}^{(a)} = \sum_{j=1}^{m} \eta_j \boldsymbol{r}^{(j)} \prod_{i=1}^{n} \langle \boldsymbol{s}_i^{(j)}, \boldsymbol{s}_i^{(a)} \rangle \tag{1.19}$$

Similarly to the linear case (1.7), the amount of evidence available to each accumulator in the model is the result of the inner product between the stimulus just presented and a general stimulus corresponding to that accumulator stored in the memory. The difference in the non-linear MLCA model is that there are now multiple inner products corresponding to the n attributes. Redefining $\mu_i^{(j,a)} = \langle \boldsymbol{s}_i^{(j)}, \boldsymbol{s}_i^{(a)} \rangle$, the previous equation can be rewritten as:

$$\boldsymbol{M}\boldsymbol{s}^{(a)} = \sum_{j=1}^{m} \eta_j \boldsymbol{r}^{(j)} \prod_{i=1}^{n} \mu_i^{(j,a)} \tag{1.20}$$

Hence, each attribute is first compared individually to the corresponding attributes of the stored stimuli. Higher evidence is obtained for similar stimuli, calculated by means of an inner product. (1.20) shows how attribute-level evidences are then multiplied to obtain the total evidence at the accumulators level. Inserting this expression into (1.14) gives the dynamics of the MLCA model upon perception of a stimulus a:

$$\boldsymbol{dx} = \left[\sum_{j=1}^{m} \boldsymbol{r}^{(j)} \eta_j \prod_{i=1}^{n} \mu_i^{(j,a)} - \boldsymbol{Dx}\right] \frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{1.21}$$

On average, the accumulator that most frequently wins the race when stimulus a is presented is the one corresponding to $\boldsymbol{r}^{(j^*)}$ such that the accumulator-level evidence is maximized. The index j^* satisfies

$$\forall j \neq j^* \in \mathcal{J}^*, \ \eta_{j^*} \prod_{i=1}^{n} \mu_i^{(j^*,a)} > \eta_j \prod_{i=1}^{n} \mu_i^{(j,a)} \tag{1.22}$$

which can be thought of as a recall function. When stimulus a is presented, the evidence for all accumulators is calculated based on the stimuli stored in the associative memory. The accumulator-level evidence is highest for the alternative that maximizes the entire product, even if other alternatives have higher attribute-level evidences for certain attributes. Importantly, the remaining alternatives play an important role in the dynamics of the MLCA model due to lateral inhibition and the stochastic nature of the accumulation. Given experimental data from stimuli with n attributes varying independently, the MLCA model not only allows for an understanding of the dynamics of evidence integration but also allows for the explicit study of the representation codes of stimuli. In Section 1.4 I discuss an example from a two-alternative task with two-attribute stimuli.

1.3.2. Multi-alternative, multi-attribute decisions

One of the consistent experimental results in the field of decision-making is the experimental finding, known as Hick's Law, that response time increases with the number of alternatives (Hick, 1952; Proctor and Schneider, 2018). Successful computational models of multi-alternative choice are expected to satisfy Hick's law (Usher and McClelland, 2001; Usher et al., 2002; Leite and Ratcliff, 2010; Brown et al., 2009; McMillen and Holmes, 2006). Because the dynamics of the MLCA are inherited from the original LCA model, the original Hick's law is satisfied, as shown by Usher and McClelland (2001). Fig 1.7 shows an example of numerical simulations with increasing number of alternatives.

However, it is not initially clear whether the use of multiple attributes should have an effect on the reaction times. According to Proctor and Schneider (2018), Hick's law is due to an increase in uncertainty by increasing the number of stimulus-response alternatives. Correspondingly, increasing the number of attributes that vary during an experiment could also increase the uncertainty. To my knowledge, no such behavioral experiment has been carried out yet.

To discover the predictions of the MLCA model I use a simple task where the correct decision depends on the first attribute, with controlled coherence, and the remaining attributes are irrelevant to the task and assigned random values. An example of such a task would be a RDK where subjects have to respond according to the direction of motion while ignoring the varying color, shape, size, etc. For m alternatives and n attributes, the matrix $\boldsymbol{M}$ is an $(m \times m^n)$ rectangular matrix that stores the associations between stimuli and responses based exclusively on the first attributes.

I simulate 1000 trials for each number of attributes under two different coherence conditions: a weak coherence of 10% and a strong coherence of 50%. Details from the simulations can be found in Appendix B. Fig. 1.8 indicates an exponential growth of response time with increasing number of attributes. As said before, there is little if any experimental data to verify this result; further experiments are needed.

1.3.3. Accounting for selective attention

The MLCA model presented in this work differs from the LCA model of Usher and McClelland (2004) in several aspects. In their proposal, in the evidence integration process, attention is restricted to a single attribute for each integration step, with some probability of switching between attributes. Usher and McClelland (2004) add a pre-input layer to their model with a node corresponding to each possible attentional focus. Such a stochastic selective attention approach is also present in MDFT (Diederich, 1997; Roe et al., 2001) and has been suggested as a crucial component of multi-attribute choice since the pioneer works of Tversky and Kahneman (Tversky, 1972; Tversky and Kahneman, 1981). Diederich and Oswald (2014) propose an attention-switching model with Ornstein-Uhlenbeck dynamics which is equivalent to that of Usher and McClelland (2004).

In MLCA, I suggest attention can be represented by a multiplicative context $\boldsymbol{c}$. Given a stimulus a, represented by vector $\boldsymbol{s^{(a)}}$, it might provide different evidence depending on attentional shifts, thus favoring different accumulators on different trials. The stimulus accompanied by its context vector $\boldsymbol{c}$ can be expressed by a tensor composition as follows:

$$\boldsymbol{c} \otimes \boldsymbol{s^{(a)}} = \boldsymbol{c} \otimes \left(\bigotimes_{i=1}^{n} \boldsymbol{s_i^{(a)}} \right) \tag{1.23}$$

A memory matrix taking contexts into account can now be written. Let $\mathcal{C}$ be the set of all possible contexts. For every context $\boldsymbol{c_h}$ in $\mathcal{C}$ there must be a stimulus $\bigotimes \boldsymbol{s_{i,h}^{(j)}}$

associated with a response $r^{(j)}$. A global associative memory $\mathcal{M}$ storing all context-dependent associations is given by

$$\mathcal{M} = \sum_j \eta_j r^{(j)} \sum_h c_h^T \otimes \left(\bigotimes_i s_{i,h}^{(j)} \right)^T \tag{1.24}$$

Assume that the attentional contexts form an orthonormal set, such that $\langle c_h, c_{h'} \rangle = \delta_{hh'}$. When a stimulus a is presented in a certain context $c_{h^*} \in \mathcal{C}$, the evidence arriving
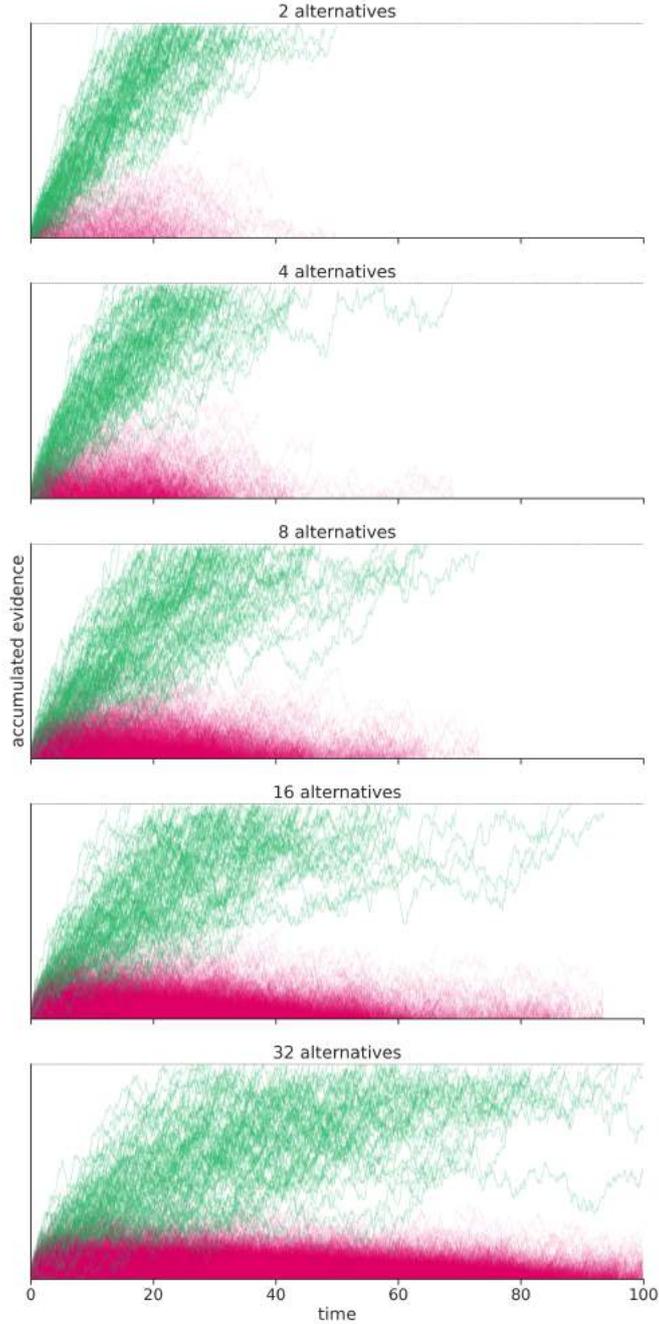
Figure 1.7: Evidence accumulation for multiple alternatives. 100 simulations are performed for each condition and their time evolutions are shown overlaid, with the correct accumulator in green and all the others in magenta. Average response time and variance increase with the number of alternatives, although the correct response is always selected.
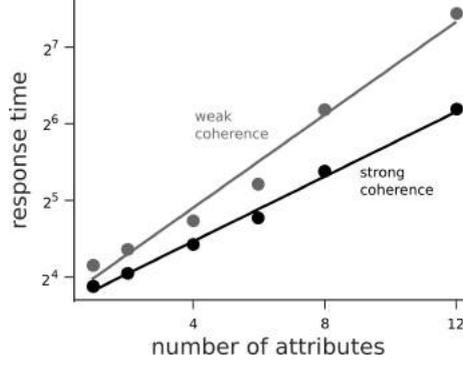
Figure 1.8: Correlations between response times and number of attributes in a two-alternative decision task. The linear fits with a logarithmic scale for the response times indicate an exponential dependency of response times with the number of attributes. The slope of the linear fit depends on the coherence of the stimuli.

at the accumulators layer is

$$\mathcal{M}\left(c_{h^*} \otimes s^{(a)}\right) = \sum_j \eta_j r^{(j)} \sum_h \delta_{hh^*} \prod_i \mu_{i,h}^{(j,a)} = \sum_j \eta_j r^{(j)} \prod_i \mu_{i,h^*}^{(j,a)} \tag{1.25}$$

which coincides with (1.20). Consequently, the inclusion of selective attention as a context allows for different evidences to be integrated by the accumulators depending on what the subjects are attending to. Instructed attention, such as cue-instructed attention on each trial, can be modeled by a single context c_{h^*}. Alternatively, stochastic attentional shifts can be captured by randomly sampling the context on a step-by-step basis. Nevertheless, the dynamics given by the MLCA model,

$$dx = \left[\mathcal{M}\left(c \otimes s^{(a)}\right) - Dx\right]\frac{dt}{\tau} + \xi\sqrt{\frac{dt}{\tau}} \tag{1.26}$$

remain as simple as before.

1.4.   An application of MLCA to cue-dependent decisions

In this Section, I test the capabilities of the model to reproduce behavioral data of a relevant experimental paradigm for multi-attribute perceptual decision-making. Mante et al. (2013) test two macaque monkeys on a color and motion RDK. The monkeys are trained to fixate a screen and wait for a contextual cue that indicates the relevant attribute for each trial: a yellow square indicates the monkeys should respond based on the motion direction of the dots (right or left), whereas a blue cross indicates responses should be based on their color (green or red) (see Fig. 1.9). The dots are presented for 750ms; during that time the monkeys integrate the noisy sensory inputs towards a choice. Following a delay of 300 to 1500ms, the monkeys make a saccade in the direction of their chosen response. The experimenters vary the motion and color coherences in both contexts, and study the monkeys' psychophysical performances by plotting psychometric curves. During the experiment, they record nearly 1500 units in the frontal eye field, a prefrontal area responsible for the selection and preparation of saccades (Schall, 2002).

Mante et al. (2013) suggest that stochastic models of selection and integration (Mazurek et al., 2003; Wang, 2002) are unable to account for context-dependent behavior. Instead, the authors use a recurrent neural network which receives color and motion sensory evidence as inputs, as well as a binary contextual cue. The network outputs a single value, integrated over time, corresponding to its choice. This recurrent neural network qualitatively reproduces the population trajectories of the recorded units in the subspace of motion, color, and choice. Nonetheless, the model fails to reproduce the psychometric curves for the irrelevant attributes, especially for monkey F who consistently fails to ignore the irrelevant attribute. Next, I show how the MLCA model presented in this Chapter can correctly account for the psychophysical behavior without contradicting the single-unit recordings of Mante et al. (2013).

Let $c_{cue} \in \{c_{motion}, c_{color}\}$ be the contextual cue, either for motion or for color, while $p_{motion} \in \{p_{100\% \ left}, \ ... \ , \ p_{100\% \ right}\}$ and $q_{color} \in \{p_{100\% \ green}, \ ... \ , \ p_{100\% \ red}\}$ are the
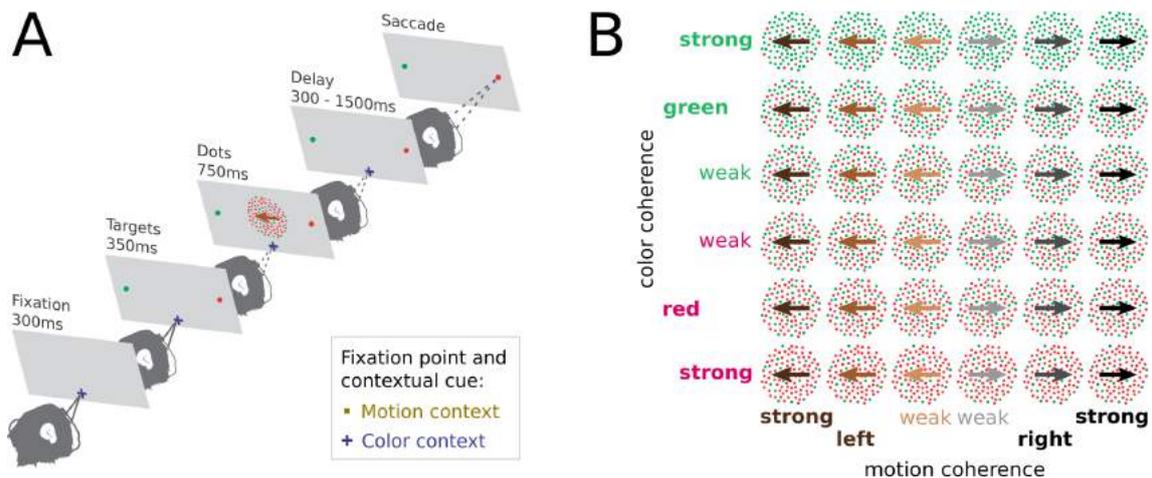
Figure 1.9: Details of the multi-attribute RDK reproduced by the MLCA model. A: Timeline. Monkeys are trained to fixate a contextual cue which indicates whether motion or color are to be attended. A RDK with preset motion and color coherences is shown for 750ms; after a random delay the fixation point disappears and the monkeys perform a saccade to the selected target. B: Stimuli. Motion coherences are set to 5, 15 and 50% moving in the defined direction, with the remaining dots moving randomly; color coherences are set to 6, 18 and 50% in the defined color, with the remaining dots colored randomly.

motion coherence and the color coherence respectively. For example, a trial with 50% *left* motion coherence and 25% *green* color coherence where the monkey should attend to motion is encoded by the input vector $\boldsymbol{c}_{motion} \otimes \boldsymbol{p}_{50\% \ left} \otimes \boldsymbol{q}_{25\% \ green}$. Given a context-dependent associative memory $\boldsymbol{\mathcal{M}}$ and a decay matrix $\boldsymbol{D}$, the dynamics of the MLCA model are given by

$$ d\boldsymbol{x} = [\boldsymbol{\mathcal{M}} \left(\boldsymbol{c}_{cue} \otimes \boldsymbol{p}_{motion} \otimes \boldsymbol{q}_{color}\right) - \boldsymbol{D}\boldsymbol{x}] \frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{1.27} $$

I suggest that the matrix memory $\boldsymbol{\mathcal{M}}$ must store the associations of the cue-dependent relevant attribute with its corresponding response. That is, with a motion cue, left-moving dots should lead to a left saccade regardless of their color, whereas with a color cue, green dots should lead to a left saccade regardless of their motion. Furthermore, given the asymmetry of these associations, left and green dots in both contexts should have stronger leftwards evidence accumulation than left and red in the motion context or right and green in the color context. This is a specific prediction given the setting of the experiment. Other settings, including alternating the position of the color responses on each trial, can be captured with an added context. A simple yet relevant simplification is to assume that the matrix memory only stores the associations for the extreme coherences (i.e. the corners in Fig. 1.9).

Defining $\boldsymbol{r}_{left}$ and $\boldsymbol{r}_{right}$ as the response codes to the left and to the right respectively, and $\eta > \eta'$ as the scaling rates for the main diagonal (i.e. left-green and right-red) and anti-diagonal (i.e. left-red and right-green) stimuli respectively, a possible context-dependent
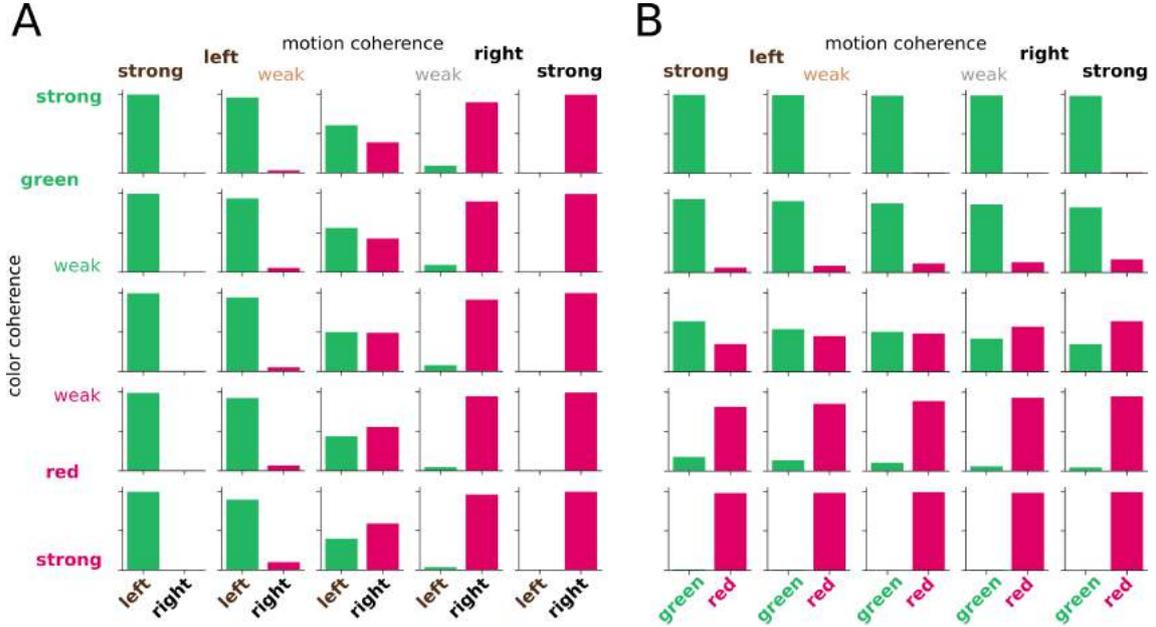
Figure 1.10: Response histograms for multi-attribute RDK simulations using the MLCA model. A: Motion context; B: color context. Each histogram corresponds to a specific motion coherence (horizontal axis) and color coherence (vertical axis), with bar heights indicating the percentage of correct responses. 1000 simulations are performed for each motion and color coherence in each context. Due to the stochastic integration, the same stimulus can entail any of the two possible responses. These results show a high dependency on the task-relevant attribute and a low dependency on the task-irrelevant attribute. However, the task-irrelevant attribute has a considerable effect on task-relevant neutral coherences (i.e. central column in A, central row in B). Mante et al. (2013) do not use completely neutral stimuli in their experiment.
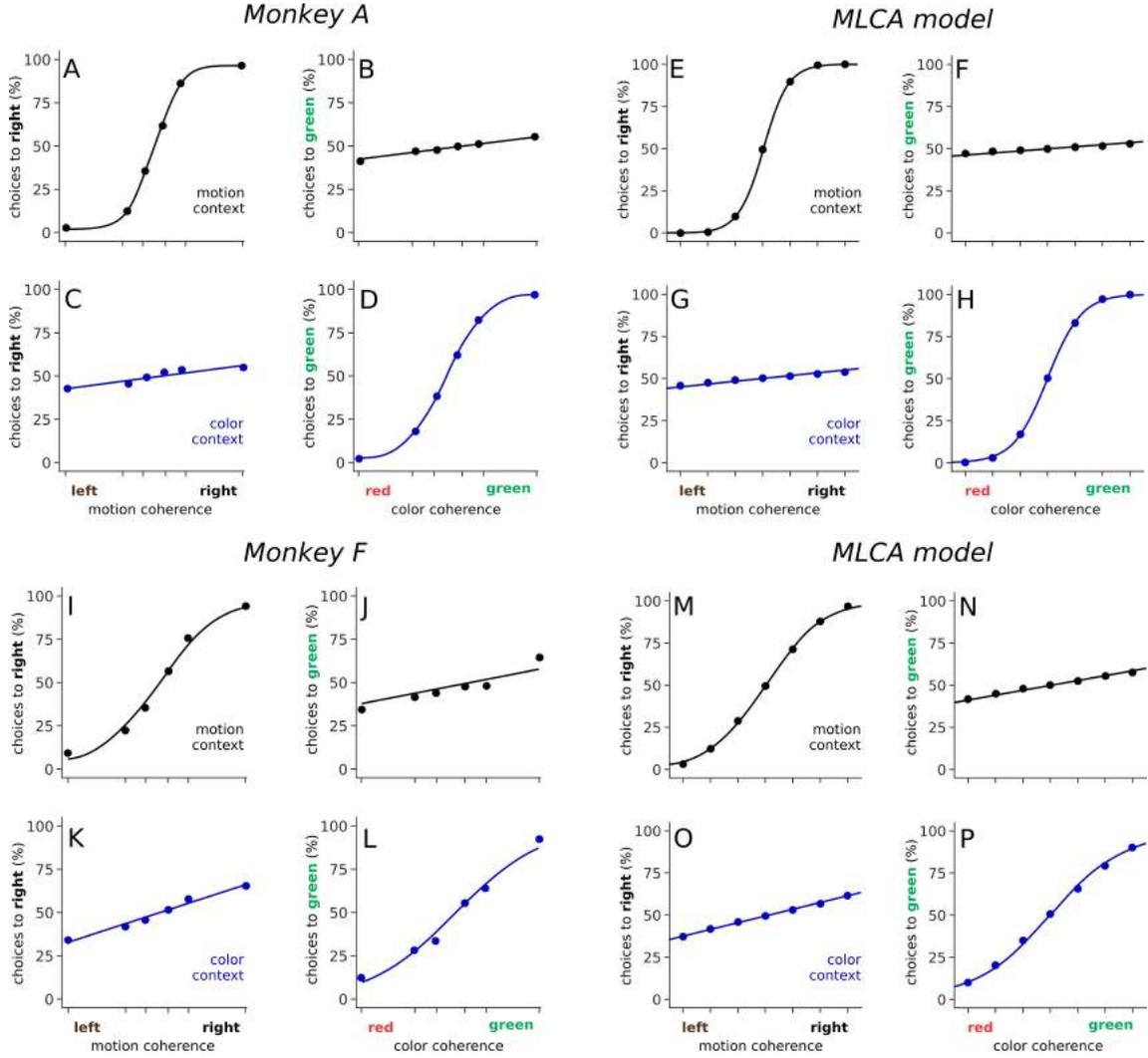
Figure 1.11: Psychometric functions for monkeys A and F and for the MLCA model. For each, psychometric curves are obtained in the motion (top) and color (bottom) contexts as functions of motion (left) and color (right) coherences. Monkey A is globally better than monkey F in discriminating the task-relevant from the task-irrelevant attribute; the MLCA model is capable of capturing this difference by means of stimuli representation vectors that can be more linearly independent (monkey A) or less (monkey F). The model shows a psychometric dependence on the task-irrelevant attribute qualitatively equivalent to that of the monkeys, an effect not shown by the recurrent neural network model of Mante et al. (2013).

associative memory for the multi-attribute RDK of Mante et al. (2013) would be

$$
\begin{aligned}
\boldsymbol{\mathcal{M}} = \eta \ & \left(\boldsymbol{r}_{left} \ \boldsymbol{c}_{motion}{}^T + \boldsymbol{r}_{left} \ \boldsymbol{c}_{color}{}^T \right) \otimes \boldsymbol{p}_{left}{}^T \otimes \boldsymbol{q}_{green}{}^T + \\
\eta' \ & \left(\boldsymbol{r}_{left} \ \boldsymbol{c}_{motion}{}^T + \boldsymbol{r}_{right} \ \boldsymbol{c}_{color}{}^T \right) \otimes \boldsymbol{p}_{left}{}^T \otimes \boldsymbol{q}_{red}{}^T + \\
\eta' \ & \left(\boldsymbol{r}_{right} \ \boldsymbol{c}_{motion}{}^T + \boldsymbol{r}_{left} \ \boldsymbol{c}_{color}{}^T \right) \otimes \boldsymbol{p}_{right}{}^T \otimes \boldsymbol{q}_{green}{}^T + \\
\eta \ & \left(\boldsymbol{r}_{right} \ \boldsymbol{c}_{motion}{}^T + \boldsymbol{r}_{right} \ \boldsymbol{c}_{color}{}^T \right) \otimes \boldsymbol{p}_{right}{}^T \otimes \boldsymbol{q}_{red}{}^T
\end{aligned}
\tag{1.28}
$$

where each line in equation (1.28) corresponds to one of the corners in Fig. 1.9. It is assumed that the matrix memory stores abstractions of the stimuli that the monkeys are exposed to during training. For example, the monkeys are trained to perform a rightward saccade upon seeing a majority of red dots in a color context, regardless of exactly how

many red dots there are and how they are moving. This learning process takes months. For our model, we are not concerned with the establishment of this memory, but rather with the fact that during the testing phase of the experiment, the associations are already stored in the monkeys' brains. Let μ be the inner product of the stored and the presented stimuli motion vectors, and ν the inner product of the stored and the presented stimuli color vectors. Upon seeing a stimulus $(\boldsymbol{p}, \boldsymbol{q})$ in the motion context, the output of the memory is given by

$$
\begin{aligned}
\boldsymbol{\mathcal{M}}(\boldsymbol{c}_{motion} \otimes \boldsymbol{p} \otimes \boldsymbol{q}) = \boldsymbol{r}_{left} \left[\eta\ \mu_{left}\ \nu_{green} + \eta'\ \mu_{left}\ \nu_{red} \right] + \\
\boldsymbol{r}_{right} \left[\eta\ \mu_{right}\ \nu_{red} + \eta'\ \mu_{right}\ \nu_{green} \right]
\end{aligned}
\tag{1.29}
$$

The bracketed terms are the readouts and therefore set the evidence in favor of the left and right responses, respectively. It can be immediately seen that if the $\boldsymbol{p}$ motion coherence is towards the left, then $\mu_{left} > \mu_{right}$ and the left motion is selected more frequently (see Fig. 1.10). Interestingly, the asymmetry imposed by setting $\eta > \eta'$ means green color coherences in the motion context have higher evidence to the leftwards accumulator than red color coherences, as seen in the monkeys' behavioral data but not in the simulations of Mante et al. (2013).

Fig. 1.11 shows how the MLCA model is capable of reproducing the psychometric curves of both monkeys. According to Mante et al., monkey F is consistently worse than monkey A. I reproduce this difference by using two different associative memories, with monkey F's stored vectors being less orthogonal (see Appendix B for details about the numerical simulations).

These results show that explicit stimuli representations and their readouts by means of context-dependent associative memories are relevant mechanisms to take into account when modeling perceptual decision-making. It should be noted that this analysis is a qualitative one. I merely show that the model put forward is capable of explaining the behavior observed in monkeys for a certain set of parameters and neural codes. A quantitative approach, not in the scope of the present work, would also lead to estimates of the parameters and restrictions on the possible neural codes.

1.5. Discussion

Perceptual decisions are based on the detection and categorization of a presented stimulus. A few hundred milliseconds after receiving the sensory information, an action is executed. Great progress has been made on the understanding of how the correct response is selected rather than any other, while simultaneously matching response times and error rates optimally (Bogacz et al., 2006). However, the success of computational models of decision-making has been hindered by their limitations when relating behavioral effects to neural dynamics. Most efforts to solve this apparent contradiction have aimed at improving the computational aspects of these models, under the assumption that single neuron recordings (e.g. Roitman and Shadlen, 2002) and accumulator nodes (e.g. Ratcliff, 1978) share a fundamental underlying mechanism.

The work presented in this Chapter points in a different direction. Rather than improving the dynamics at the computational level, focus is transferred towards the representational level. Although using evidence accumulation nodes (Usher and McClelland, 2001), I suggest these are modeling mesoscopic levels of activity at decision-making regions in the brain and need not behave analogously to single neurons. Attention is rather shifted to the transmission of information from upstream sensory areas into the decision-making region. One such example is the association of stimulus representation activity in MT and its readout and accumulation in LIP (Mazurek et al., 2003).

The associations between representations and readouts are modeled by associative memory matrices: linear associations for single-attribute stimuli represented by a single encoding vector and tensor product associations for multi-attribute stimuli. Compared to other decision-making models, including neurally-inspired sequential sampling ones, the model presented here offers the advantage of a biologically plausible mechanism, which the brain could employ when transferring information between different processing levels. This proves particularly relevant for capturing context-dependent effects due to multi-attribute stimuli. The results from Section 1.4 show that the model is uniquely capable of reproducing the psychometric functions for two different monkeys in a multi-attribute RDK experiment conducted by Mante et al. (2013). Although the authors of that study state that their interest is to reproduce single-cell recordings rather than behavior, and they use a recurrent neural network to do so, it is worth mentioning that their model and the MLCA are not mutually exclusive. As mentioned previously, sequential sampling models should be thought of as mesoscopic models that can merely account for the global dynamics of entire populations of neurons.

A key innovation introduced in this thesis is the vector-matrix representation for the LCA and DDM models (Section 1.2.1) and their multi-attribute extensions (Section 1.3). The vectorial nature of the variables comes from the fact that all neurocognitive activity corresponds to patterns of activity of large groups of neurons, whose natural mathematical representations are vectors. In addition, by using vectors instead of scalar variables, the mathematical formalism presented here is readily generalized to model multi-attribute, multi-alternative decision-making.

The relative simplicity of the algebraic LCA equation is due to the readout function used: a linear associator between the single-attribute stimuli representation nodes and the accumulator nodes. This is an ad hoc assumption in the present work since it is the simplest and frugal neural representation for input-output relationships. Similarly, the readouts from multi-attribute representations are modeled with context-dependent associative memories. These are vector state models that use tensor product variable binding, which demonstrated representational power and the possibility to develop theoretical neurocognitive approaches using the well-known capacities of linear algebra.

Whether these are the exact operations being performed at the neural substrate will only be known as technological developments allow for more precise neurophysiological recordings. Nonetheless, as the behavioral-level results of our model match perfectly the experimental data of Mante et al. (2013)., I suggest a context-dependent associative memory can be an adequate first approximation to the real multi-attribute evidence integration mechanism.

As a final comment, let me bring notice to the fact that the use of associative memories throughout this thesis could lead to unwanted misunderstandings related to perceptual learning. Much has been said about perceptual learning which is well beyond the scope of the present work (e.g. Goldstone, 1998). My use of the term associative memory refers exclusively to a category of models of information transmission between successive layers in a natural or artificial neural network. Regarding perceptual decision-making tasks such as making an eye movement in the direction of moving dots, the associative memory is a simple model that relates the perception of the dots to the corresponding action. As trivial as it might seem, some type of learning is necessary to establish that stimulus-response association rather than, for example, making an eye movement in the direction opposite to the moving dots. The process of perceptual learning may be an extremely complex lifelong one (Gilbert et al., 2001). When subjects perform a perceptual task, the associations between the stimuli and the responses are already established, and it is partly the goal of the experimenter to discover them. The MLCA model intends to facilitate that discovery.

This concludes the first Chapter of this thesis, which serves as the mathematical foundation for the following. The MLCA model developed and tested up to this point is next used to reproduce two other well-studied experimental observations: the Simon effect and the SNARC effect. Both these effects share with the multi-attribute RDK of Mante et al. (2013) the influence of task-irrelevant information on psychometric results.

Chapter 2

Spatial associations of stimuli and responses

In this Chapter, I include spatial representations into the MLCA model presented in Chapter 1 in order to reproduce the Simon effect. Section 2.1 gives an introduction to the idea of spatially represented stimuli and responses in the brain. In Section 2.2, I look at stimulus-response compatibility effects. Incongruencies between certain attributes of stimuli and responses are related with higher response times and error rates. Incongruencies due to the locations in space are known as the Simon effect, which is discussed in detail in Section 2.3. After presenting the main experimental findings of the Simon effect, I introduce the computational models that have been proposed to explain it. Finally, in Section 2.4 I present two unsuccessful models followed by a successful dual-route MLCA account of the Simon effect.

2.1. The representation of space in the brain

The idea of a mental representation of space dates back to the preludes of the cognitive revolution. In his classical experiment from 1946, Edward C. Tolman trained a group of rats to find food following a curved path in an elongated apparatus. When the apparatus was changed for another one containing a number of different possible paths, rats chose the path that pointed towards the original location of the food, disregarding the shape of the original path, which led the authors to conclude that rats were able to represent the spatial location of their specific goal rather than following the operant conditioning strategy of repeating their learned behavior (Tolman et al., 1946). In fact, rodents have been a particularly fruitful resource for discoveries about spatial representation and navigation. Cheng (1986) found that rats use geometric and non-geometric visual information to represent the space around them. O'Keefe and Dostrovsky (1971) noticed that certain cells in the hippocampus show activity only when rats are situated in a particular part of a testing platform; they called them *place cells*. Hafting et al. (2005) found cells in the entorhinal cortex of rats that fire at specific regions which form a mesh of triangles that range over the entire available space; they called them *grid cells*. In 2014, the Nobel Prize in Physiology or Medicine was awarded to John O'Keefe, for the discovery of place cells, and to May-Britt and Edvard Moser, for the discovery of grid cells. Other notable types of neurons which contribute to the representation of space include boundary cells (Barry et al., 2006), head direction cells (Taube et al., 1990), and speed cells (Kropff et al., 2015). All these are believed to play a fundamental role in what Tolman called a cognitive map.

The discovery of these specialized types of neurons has elevated the importance of spatial cognition. The hippocampus, in particular, is the main region of interest regarding
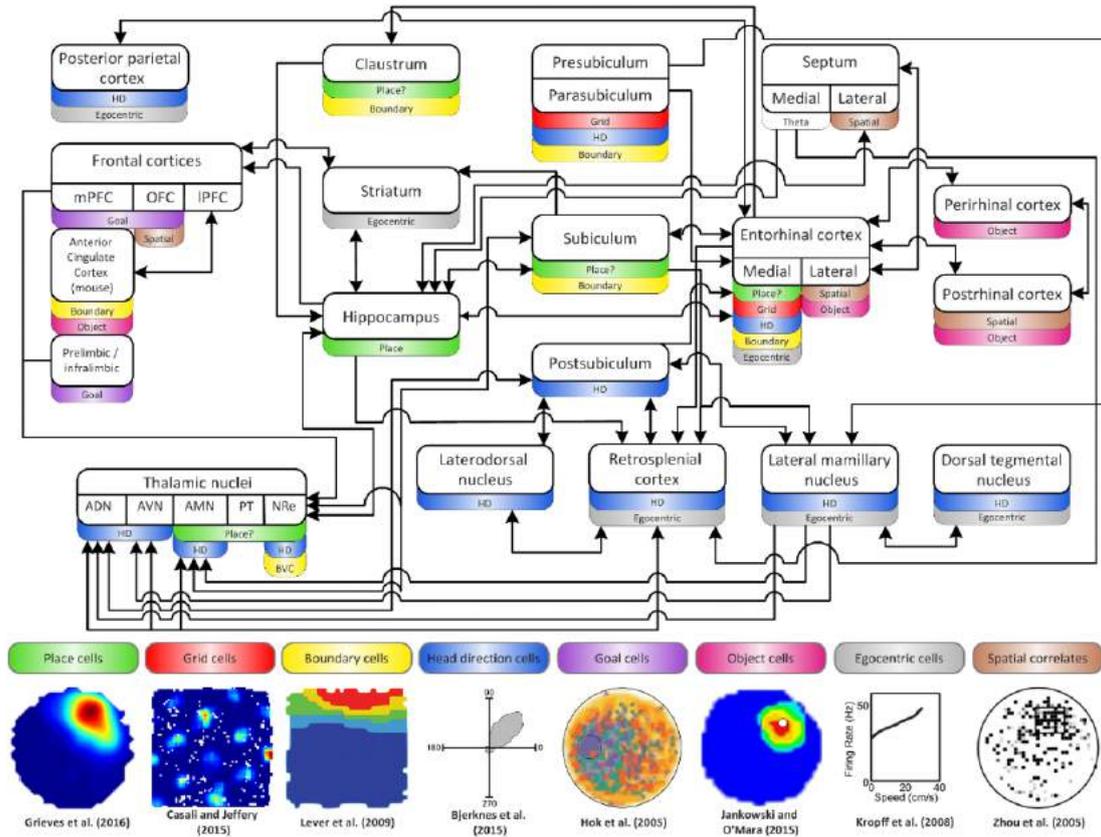
Figure 2.1: Schematic overview of spatial representation in the brain (Grieves and Jeffery, 2017). The focus throughout this thesis is set on the IPS in the PPC, which has head direction cells and egocentric cells (shown in the diagram) as well as retina-centered neurons corresponding to the visual system (not shown).

the establishment of new, spatially represented memories (Hartley et al., 2014). The current understanding of the representation of space in humans and animals involves an extensive network of cortical and subcortical areas, such as the hippocampus, the entorhinal cortex, the posterior parietal cortex, the medial prefrontal cortex and the orbitofrontal cortex, among others (Grieves and Jeffery, 2017; Fig. 2.1). Here, I am not interested in how spatial information is used for navigation, but rather how the perception of stimuli in different positions changes the accuracy and speed of responses. I focus mainly on the parietal areas involved in the representation of space.

The fact that spatial information is somehow encoded in the parietal cortex is well-known. Mishkin and Ungerleider (1982) performed lesions to different regions and connections in the brains of monkeys and suggested the existence of two general pathways for the processing of incoming visual stimuli: a ventral pathway and a dorsal pathway. The former, known as the "what pathway", includes areas of the visual cortex V1, V2 and V4, and the inferotemporal areas of the temporal lobe, and deals with semantic processing, such as identifying objects or faces. The latter, known as the "where pathway", connects areas V1, V2, and V3 with area MT and the posterior parietal cortex (PPC), and deals with locations and motion. The idea of two separate pathways may be an oversimplification, since there is evidence of strong connections between both pathways, but it is nonetheless a useful model of the brain in that it stimulates innovative research (McIntosh and Schenk, 2009).

The PPC is not restricted to the spatial coding of visual information. Physiological
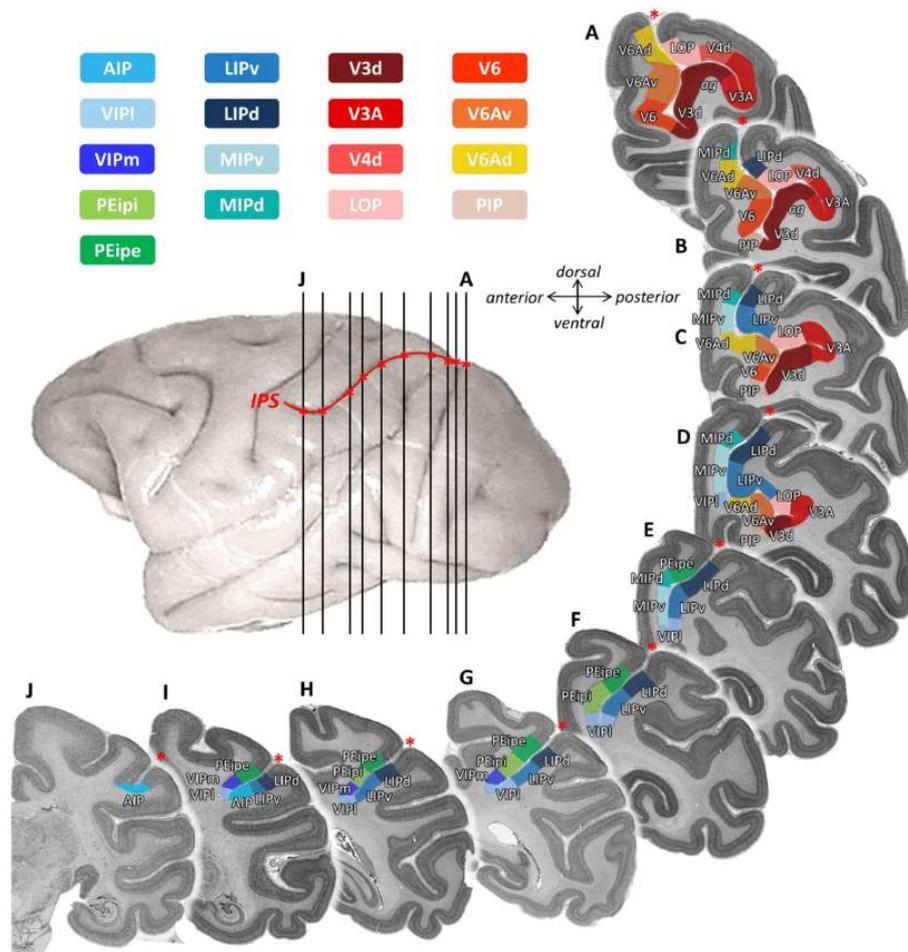
Figure 2.2: Anatomical topography of IPS areas in the left hemisphere of a macaque brain (Niu et al., 2020). The main areas discussed here are AIP, VIP, LIP and MIP.

studies with monkeys show a shared multisensory representation of space for vision, hearing (Andersen et al., 1997) and touch (Azañón et al., 2010). Similar results are obtained with transcranial magnetic stimulation in humans (Lewald et al., 2002). Andersen et al. (1997) suggest the PPC deals with coordinate transformations, such as the conversion of retinotopic visual information into oculocentric coordinates, which allow both humans and non-human animals to make saccades towards a target. Such conversions require not only visual perception, but also proprioceptive and vestibular information, as well as connections with executive and premotor areas. The associative PPC is both anatomically and functionally perfectly suited for this role (Paré and Dorris, 2011). A model by Beck et al. (2011) indicates that an object's position can be coded by combining the retinotopic position of the object in the visual field with the proprioceptive position of the eyes. This would result in an abstract spatial representation which could be converted into head-centered, body-centered, or even world-centered coordinates (Andersen et al., 1997).

The IPS is the main sulcus on the lateral PPC. Several sub-areas of the IPS have been identified as having specific roles for representing spatial information in different reference frames. Of course, there are anatomical differences between humans and monkeys, but physiological and imaging studies have allowed researchers to find putative functionally equivalent areas between both species (Grefkes and Fink, 2005). Here, I refer to the sub-areas of the IPS by the nomenclature used for the monkey brain, as shown in Fig. 2.2.

The LIP has retina-centered neurons with receptive fields similar to those of the visual

cortex, so that both moving the stimulus or moving the eye alter which neurons respond and how. LIP is one of the areas where evidence accumulation can be detected (see Section 1.1.1). Colby and Goldberg (1999) highlight the role of attention for neurons in LIP: firing rates depend on whether the monkey is attending to the stimulus or not, such as remembering the position of a stimulus.

The MIP also has retina-centered neurons with receptive fields. Unlike LIP, area MIP is mainly concerned with the preparation of a motor action towards a target. Neurons in MIP also receive information about the hands positions and movements; they are believed to transform a target's position into a movement vector (Cohen and Andersen, 2002), thus being responsible for the coordination of hands and eyes during reach movements.

The ventral intraparietal area (VIP) has head-centered neurons with broader receptive fields. If a monkey's head is kept fixed during a task involving saccades, the receptive fields of area VIP do not change. VIP neurons are believed to be responsible for the multisensory processing of visual, auditory and tactile information (Duhamel et al., 1998), as well as coding for head position and movement (Klam and Graf, 2003). Area VIP is mainly concerned with the perception of the body itself and objects in close proximity.

The anterior intraparietal area (AIP) has head-centered neurons. Area AIP combines visual and tactile information during the preparation of a hand movement. AIP neurons are particularly active when reaching and grasping 3D objects. They are responsive to size, shape and orientation (Murata et al., 2000). Area AIP has strong connections with the ventral premotor cortex.

Ultimately, there are several different spatial representations in IPS, depending on the characteristics of the task. Areas LIP and VIP are mainly involved in stimuli perception, with the former responsible for visual tasks and keeping in working memory the position of stimuli, and the latter responsible for coding the positions of objects in close proximity. Areas MIP and AIP are responsible for the preparation of reaching towards a target and grasping an object, respectively.

It is also worth mentioning that the spatial representations of both stimuli and actions appear to be discrete and categorical rather than continuous. Huttenlocher et al. (1991) conduct an experiment in which subjects have to report the location of a dot inside a circle from memory. They find a high bias towards certain directions and distances from the center, as if subjects were dividing the circle into regions and estimating the dot's location at the centers of those regions.

2.2. Stimulus-response compatibility effects

On January 8, 1989, a British Midland Airway's Boeing 737-400 crashed during a flight between London and Belfast, resulting in 47 casualties and 74 injured. The accident was caused by a faulty left engine, which caught fire and filled the aircraft with smoke (Learmount and Norris, 1990). The pilots, who were unable to see where the smoke was coming from, had to make a quick decision relying exclusively on the dashboard instruments. The engine displays were grouped into two groups. On the left were the primary instruments, with the left engine ones in the left column and the right engine ones in the right column. The secondary instruments were displayed in a similar fashion to the right. The throttles were placed directly beneath each of the two groups of engines. The pilots had a combined 1000 hours of flying Boeing 737s, but merely 76 hours with the 737-400, and therefore were unable to interpret the displays correctly and quickly enough. When the left engine showed signs of trouble through the secondary instruments, which were displayed over the right engine throttle, the pilots shut down the right engine, thus leaving the aircraft adrift (Proctor and Vu, 2006). They were unable to recognize the mistake in time. The Kegworth disaster, as the accident was named, is an example of incompatibility between stimuli and response codes.

Stimulus-response compatibility (SRC) effects correspond to psychometric differences in response times and error rates between situations in which either the stimuli, the responses, or both, are modified. SRC comes from the mapping of stimuli and responses into a common representation, such that high compatibility results in natural associations and low compatibility results in unnatural ones (Proctor and Vu, 2006). Research on SRC began in the 1950s, with the work of Paul Fitts and colleagues (Fitts and Seeger, 1953; Fitts and Deininger, 1954). In his own words, "A task involves compatible stimulus-response relations to the extent that the ensemble of stimulus and response combinations comprising the task

Figure 2.3: SRC effect example (Fitts and Seeger, 1953). A light is shown at one of the positions of the stimulus sets S_A or at one or two of the positions of S_B and S_C. Responses are made by moving a single level with the response sets R_A and R_B or two levers with R_C. Response times and errors increase due to incompatibilities between stimuli and response sets.

results in a high rate of information transfer. [...] This interpretation makes use of the idea of a hypothetical process of information transformation or recoding in the course of a perceptual-motor activity, as assumes a degree of compatibility when recoding processes are at a minimum" (Fitts and Seeger, 1953, p. 199).

Fitts' work focused on the reaction times of human subjects for different types of displays of stimuli and responses. For example, Fitts and Seeger (1953) use 3 different stimuli sets, which indicate a position on one of 8 possible radial positions, and 3 different response sets, which allow subjects to respond towards those directions (Fig. 2.3). The matching stimulus and response displays have the lowest reaction times, in agreement with Fitts' proposal that incompatible combinations take longer because an additional transformation is necessary.

SRC effects can be classified according to several different characteristics. Firstly, SRC can occur at a set level or at an element level. Set-level compatibility refers to different pairings of stimuli and responses. For example, the response set of written words {*left*, *right*} is more compatible with the set of spoken words {/*left*/, /*right*/} than with the set of arrows {←, →}. However, the set of arrows is more compatible than the set of written words if the response consists of making a saccade to the left or to the right (Wang and Proctor, 1996). Element-level compatibility involves different pairings of stimuli and response elements within fixed sets. For example, if the sets are the written words {*left*, *right*} and the arrows {←, →}, there is higher compatibility when the word *left* is paired with the arrow ← than with the arrow →.

These examples highlight the idea that SRC effects are the result of natural, or unnatural, associations between stimuli and responses. Kornblum et al. (1990) propose a theoretical model of SRC based on what they call the dimensional overlap between stimuli and responses: "We define dimensional overlap as the degree to which two sets of items are physically or conceptually similar" (Kornblum, 1991, p. 5). These physical or conceptual similarities can be hardwired into the brain, culturally acquired over development, or result from short-term associations in working memory. For example, a sequence of ordered letters {*A*, *B*, *C*} has high similarity with the ordered set of numbers {*1*, *2*, *3*}. Kornblum et al. (1990) also propose a taxonomy of the different types of tasks

Table 2.1: A taxonomy of stimulus-response compatibility ensemble types (Kornblum, 1994).

Type	S-R overlapping dimensions (relevant)	S-R overlapping dimensions (irrelevant)	S-S overlapping dimensions	Representative study
1	no	no	no	Most experimental control conditions, e.g. in Fitts and Deininger (1954)
2	yes	no	no	Fitts and Seeger (1953).
3	no	yes	no	Simon (1969); Wallace (1971).
4	no	no	yes	Eriksen and Eriksen (1974).
5	yes	yes	no	Hedge and Marsh (1975).
6	yes	no	yes	(None).
7	no	yes	yes	Kornblum (1994).
8	yes	yes	yes	Stroop (1935); Simon and Rudell (1967)

based on SRC dimensional overlap. The authors classify tasks by whether the stimuli, which might be composed of both task-relevant and task-irrelevant attributes, and the responses have dimensional overlap. A total of 8 ensemble types are identified (see Table 2.1).

Here, I focus exclusively on the ensemble type 3, for which dimensional overlap can only be identified between the response set and the irrelevant stimulus dimension. When said dimension is a spatial one, the SRC effect is called the Simon effect.

2.3. The Simon effect

The Simon effect was first reported by, and named after, J.R. Simon. In a series of experiments from the late 1960s, Simon and colleagues studied a SRC effect where subjects had to press one of two lateralized buttons in response to a visual or auditory stimulus, such as a color or a tone. The stimulus was also lateralized, so that there was a dimensional overlap between the task-irrelevant position and the response (Simon and Small, 1969; Simon, 1969; Simon and Craft, 1970). In Simon and Small (1969), they presented subjects with a high-pitch tone and a low-pitch tone, and subjects had to press either a left button or a right button for each one. On congruent trials, i.e. when the sound was presented to the ear on the same side as the button corresponding to that tone, responses were faster than on incongruent trials. The Simon effect consists of larger response times and error rates due to a task-irrelevant spatial coding of the stimulus.

Some authors claim that the first report of a Simon effect comes from Simon and Rudell (1967). In that experiment, subjects heard the words *left* and *right* instead of tones. According to the taxonomy of Kornblum (1994), this adds both a relevant stimulus-response and a stimulus-stimulus dimensional overlap; in this sense, the effect reported by Simon and Rudell (1967) is sometimes termed the spatial-Stroop effect (Hommel, 2011). Fig. 2.5 shows a schematic interpretation of the different overlapping dimensions for the Simon effect, the spatial-Stroop effect, the Stroop effect, and the flankers effect.

2.3.1. Experimental results

Following the results of Simon and Small (1969), Simon (1969) investigated whether the faster response times are simply due to the lateralized cerebral hemispheres. He repeated the original experiment but with unimanual responses, by using a lever with leftwards or rightwards movements, instead of two buttons. The effect was still present, so the two hemispheres could not be the source of the effect. Wallace (1971) went a step further and proved that the effect can not be attributed to an anatomical origin. He repeated the Simon task but asked subjects to cross their hands, so that the *right* hand was responsible for the *left* response, and *vice versa*. He reported a Simon effect with respect to the response location, not the hand that executes the action.

Since its discovery, the Simon effect has also been reported with tactile stimuli (Hasbroucq and Guiard, 1992) and several types of visual stimuli, such as colors (Simon, 1990), geometric shapes (Nicoletti and Umiltá, 1989), letters (Proctor and Lu, 1994), and
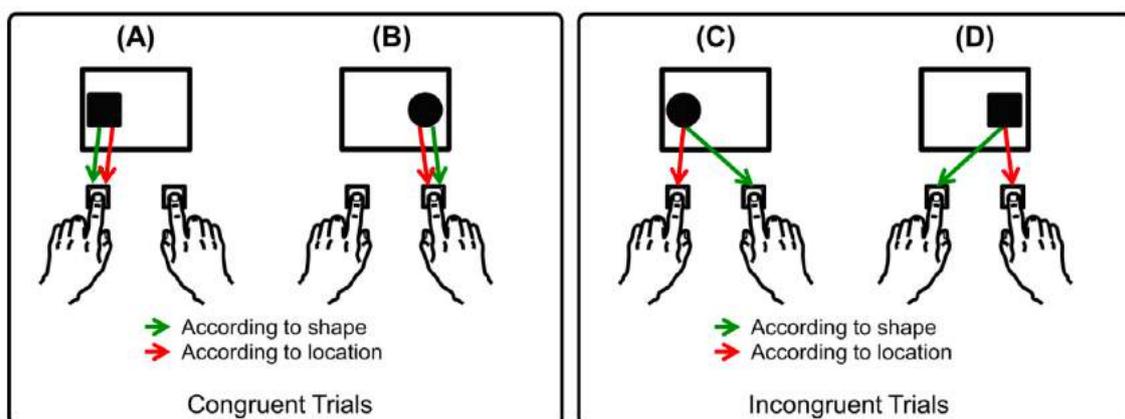
Figure 2.4: Congruent and incongruent trials in a Simon task (Stocco et al., 2017). The task-relevant shapes are arbitrarily assigned one button each; locations interfere in the response-selection process even though they are task-irrelevant.
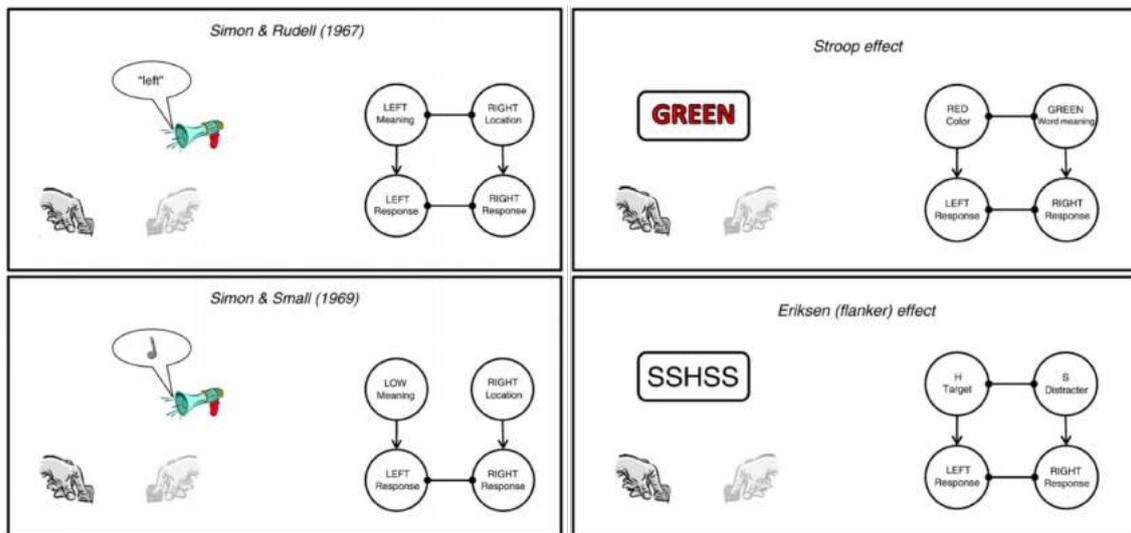
Figure 2.5: Dimensional overlap in Simon and related effects (adapted from Hommel, 2011). Only the Simon effect (Simon and Small, 1969) can be attributed to incongruencies at the response-selection stage; the other three effects also include dimensional overlap between the task-relevant and task-irrelevant attributes of the stimuli.

light intensity (Proctor et al., 1992). Mewaldt et al. (1980) find a Simon effect with accessory stimuli that are not involved in the task, namely lateralized sounds during a visual discrimination task with centralized stimuli. The effect has also been reported in pigeons (Urcuioli et al., 2005) and rats (Courtière et al., 2007), and a similar phenomenon was also observed in monkeys (Georgopoulos et al., 1989). All these experiments report higher response times for incongruent trials.

Spatial codes

Because the Simon effect is still present when subjects cross their hands and when accessory stimuli are presented simultaneously, the leading theory is that stimuli are automatically represented spatially, even though this is irrelevant for the task. This spatial code of the stimulus may then interfere with the spatial code of the response, as predicted by Wallace (1971). Strong connections between different areas of the IPS may be responsible for this interference.

It is unclear why the spatial code for a stimulus, although irrelevant, is formed. According to the premotor theory of Rizzolatti et al. (1987), by attending *to the left*, a premotor command to move the eyes towards the left is formed, but not executed. Nicoletti and Umiltá (1989) attribute the generation of spatial codes to attention, so that the codes for *left* and *right* are defined with respect to the attentional focus. They design a Simon-like experiment with distractors, such that subjects are unable to give their full attention to the stimuli, and find that the Simon effect disappears. Lamberts et al. (1992) find SRC effects in three different reference frames: hemispace, i.e. the absolute position of a stimulus with respect to the world; visual hemifield, i.e. the position of the stimulus with respect to the fovea; and relative position, i.e. the position of the stimulus with respect to the other positions in which it can appear.

Delta plots

An important breakthrough in the understanding of the Simon effect came following a distributional analysis of response times. De Jong et al. (1994) perform a technique
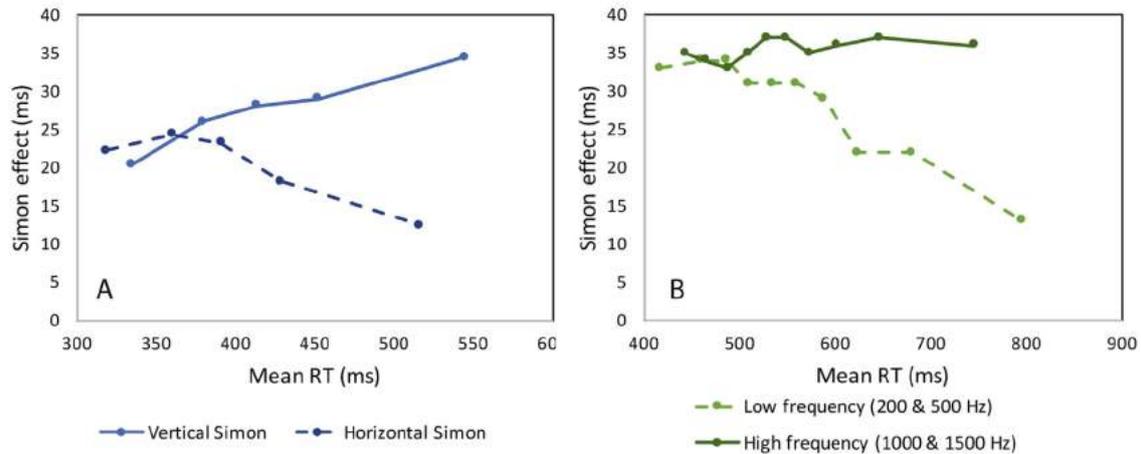
Figure 2.6: Delta plots for different Simon tasks (Salzer et al., 2017). A: visual Simon tasks; B: auditory Simon tasks. Positive slopes indicate an effect which increases with response time; negative slopes indicate an effect which decreases with response time. The different slopes within Simon tasks are attributed to the differences in the automatic process of a dual-route model (De Jong et al., 1994).

called delta plots, which consists of splitting the response times for the congruent and incongruent conditions into bins, and plotting the time differences between corresponding bins against their mean response time. Their results indicate several interesting insights about this effect. First, they obtain negative slopes, indicating that the Simon effect disappears or can even be reversed for long response times. The same can be observed by artificially increasing response time, such as increasing the difficulty of discriminating stimuli (Lammertyn et al., 2007). Second, when they test different types of Simon-like effects, De Jong et al. (1994) find varying slopes in the delta plots.

Delta plots have since become a standard procedure for studying the Simon effect, and also one of the hardest experimental results to explain. Salzer et al. (2017) review a number of different experiments across sensory modalities (Fig. 2.6). The differences in the delta plots between tasks can only be explained if the underlying response time distributions for congruent and incongruent conditions are different (Pratte et al., 2010). Positive delta plot slopes, i.e. an increasing Simon effect with increasing response time, are attributed to a congruent distribution with small variance and an incongruent distribution with large variance. The opposite is true for negative delta plot slopes.

Importantly, sequential sampling models with fixed integration noise, such as the DDM and LCA introduced in Chapter 1, cannot produce negative delta plot slopes (Salzer et al., 2017). If the mean of the response time distribution is larger, i.e. slower response times, the variance of the distribution is correspondingly larger. A possible solution is to consider a dual-route.

Dual-route model

The dual-route model, originally suggested by Kornblum et al. (1990) and later by De Jong et al. (1994), consists of two independent pathways for the processing of stimuli in a Simon-like task. First, an automatic or fast route processes the stimulus' position and prepares the motor program to respond towards it. Because this is a task-irrelevant attribute, the automatic route is inhibited and the motor program is paused. A second, intentional and slower route deals with the cognitive processing of the stimulus itself; once it is recognized, the associated response is prepared. In a congruent trial both responses are the same, and therefore the original motor program can be executed; in an incongruent trial, the original
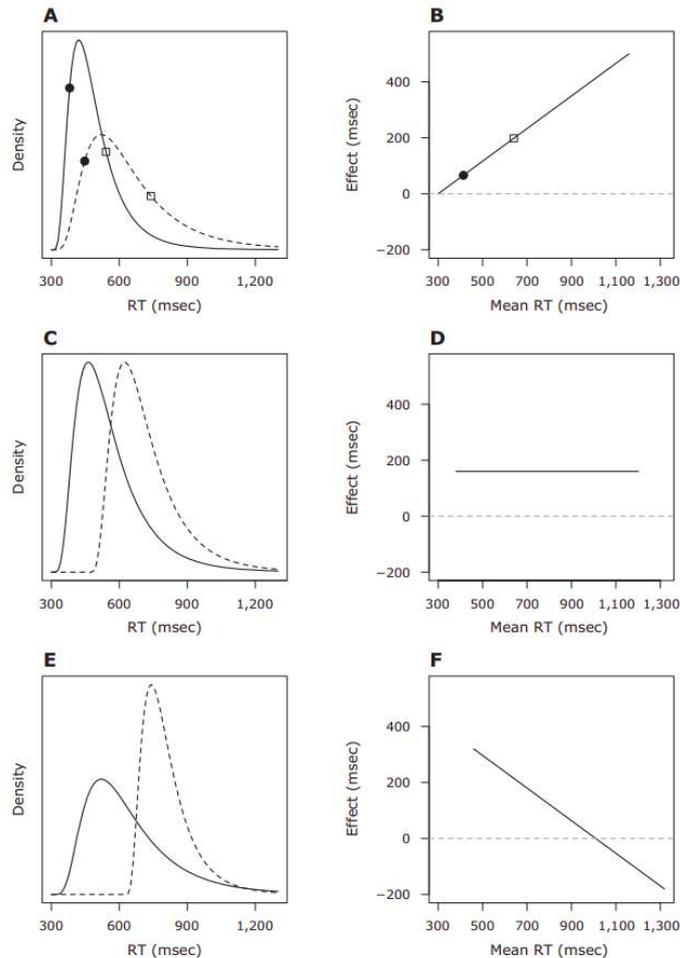
Figure 2.7: Response times distributions and their resulting delta plots (Pratte et al., 2010), with congruent conditions with filled lines and incongruent ones with dashed lines. In all three cases, the average response time for a congruent condition is lower than that of an incongruent condition. However, negative slopes in delta plots are only possible if the standard deviation of response times for a congruent condition is larger.

motor program must be aborted and the correct one must be prepared and executed, taking longer time than in the congruent condition. For longer response times, the original motor program will have been paused for more time, and it will need to be prepared again both in congruent and incongruent trials, which might explain the negative slopes in delta plots.

The dual-route model is anatomically consistent. If the decision of the program to be executed is made in the IPS, then the dorsal pathway has fewer and faster connections than the ventral pathway. The location of the stimulus can therefore influence the decision-making process earlier than the content of the stimulus itself.

Georgopoulos et al. (1989) conduct a physiological study with monkeys that reveals that the premotor cortex of the hemisphere contralateral to the stimulus' position is activated shortly after presentation, even though no response is executed. Similarly, electromyographic activity in humans during a Simon task shows early muscular activity, although again no response is executed (Servant et al., 2015). Using a technique called lateralized readiness potentials (LRPs), which consists of taking the difference between opposite electrodes in motor areas after an event-related potential EEG experiment, researchers can find differences in neural activity between congruent and incongruent trials. LRPs consistently show asymmetric activation during the preparation of a

response for SRC effects (Eimer, 1997; Masaki et al., 2004) and the Simon effect (Valle-Inclán, 1996; Wiegand and Wascher, 2005).

2.3.2. Computational models

Several theoretical accounts have been suggested in an attempt to explain the Simon effect. Here, I discuss some of the main mathematical and computational models that have tried to reproduce it. Some of the more conceptual explanations that I do not detail include the proposed architecture for the dual-route model (Kornblum et al., 1990) and the theory of event coding (Hommel et al., 2001).

An algorithmic model

Paul Rosenbloom and Allen Newell were the first to propose a symbolic information processing model to account for SRC effects, and particularly the Simon effect (Rosenbloom, 1986; Rosenbloom and Newell, 1988). Their model consists of an algorithm where response times correspond to the number of operations needed to transform the stimulus into a response. An example of the algorithmic model applied to the experiment of Fitts and Seeger (1953) with stimulus set S_B and response set R_A in pseudocode is

0. BEGIN 1. FIND STIMULUS WHERE LIGHT ON 2. IF (STIMULUS IS ONE LIGHT): 3. GET ANGLE OF STIMULUS 4. PUSH LEVER WITH ANGLE 5. ELSE IF (STIMULUS IS TWO LIGHTS): 6. GET ANGLE1 OF STIMULUS FIRST LIGHT 7. GET ANGLE2 OF STIMULUS SECOND LIGHT 8. SET ANGLE AS AVERAGE OF ANGLE1 AND ANGLE2 9. PUSH LEVER WITH ANGLE 10. END

The algorithm begins by getting the value of the angle to the light. If there is only one light, then the response can be executed for that exact angle. If there are two lights, then the angle for that second light is obtained, the average of the two is calculated, and that is the angle used for the response. During a congruent trial the response is executed at the first conditional statement, whereas during an incongruent one the response is executed after the second conditional statement. The algorithmic model corresponds to the intuition of Fitts and Seeger (1953) that larger response times are due to more recoding being needed. However, this symbolic information processing account cannot explain other behavioral and neural findings.

Connectionist models

Several connectionist models have been proposed to explain the Simon effect. Zorzi and Umiltá (1995) use a two-layer architecture with two position nodes and two non-spatial attribute nodes that feed into two response nodes with lateral inhibition (Fig. 2.8). The response nodes integrate inputs until a threshold. When a trial begins, one of the location nodes and one of the attribute nodes are activated. If they feed the same response node, i.e. a congruent trial, then that response node is selected. If they feed different response nodes, i.e. an incongruent trial, then response time is slower. Importantly, their model includes two time-dependent parameters. First, the spatial codes decay shortly after being formed, as suggested by Hommel (1994), so that the activation of the location nodes lasts only a fraction of the trial. Second, the activation of the attribute nodes is delayed as a representation of task difficulty, so that harder discrimination tasks lead to slower response times (see Lammertyn et al., 2007). This model is fairly simple but it does obtain response times quantitatively equivalent to experimental ones.

Tagliabue et al. (2000) extend the original model from Zorzi and Umiltá (1995) with short-term and long-term associations. Long-term ones are fixed prior to the experiment, whereas short-term ones can be altered during the experiment. By modifying the
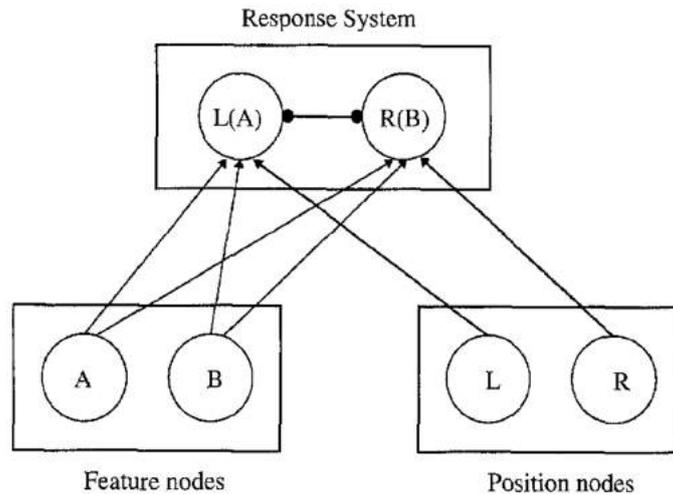
Figure 2.8: Connectionist model of the Simon effect (Zorzi and Umiltá, 1995). Independent feature and location nodes feed competing response nodes, which accumulate the input until one of them reaches a threshold.

connections of the model from trial to trial they are able to explain a reduction of Simon effect that comes from practice (Rottermann and Vu, 2009).

A third relevant connectionist account is the parallel distributed processing model of Zhang et al. (1999). They propose a three-layer architecture that intends to represent the original idea of a dual-route model by Kornblum et al. (1990). They include an intermediate layer that receives as input the two attributes of the stimulus shown and outputs a response code which is integrated until a threshold. This model achieves qualitatively satisfactory results but, as all previous connectionist models, fails to account for the distributional analysis of De Jong et al. (1994).

A cognitive architecture

Stocco et al. (2017) suggest the Simon effect can be explained by a dual-route model, where the processing of task-irrelevant information, such as the position of the stimulus, is modulated by the direct and indirect pathways of the basal ganglia. To test this hypothesis they implement a cognitive ACT-R architecture which includes modules for the stimuli representations, working memory, and long-term memory. Because ACT-R does not include a module for the two pathways of the basal ganglia, Stocco et al. (2017) use two separate productions. For example, the direct pathway may favor the processing of location while the indirect pathway prevents it from being kept in working memory. Their results show that a higher role for the indirect pathway of the basal ganglia leads to an easier inhibition of irrelevant spatial information during incongruent trials.

Sequential sampling models

The final category of computational models used to reproduce the Simon effect are sequential sampling ones. As explained before, a single stochastic accumulation process cannot account for the variety in delta plots observed across different types of Simon-like experiments (Salzer et al., 2017). Two different proposals have been given.

Servant et al. (2015) propose a model with two phases based on the LCA (Usher and McClelland, 2001). During the first phase, the input to the accumulators consists of both the location of the stimulus and its non-spatial attribute. During the second phase, only the non-spatial attribute serves as input. The focus of their work is to find partial errors, i.e.
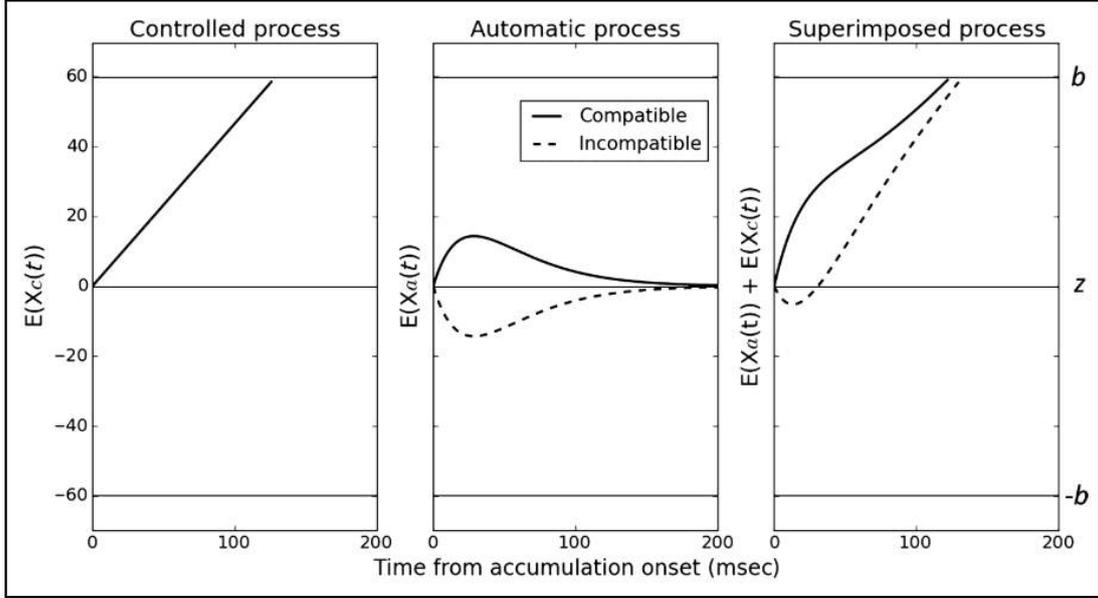
Figure 2.9: Dual-route model of the Simon effect (Servant et al., 2016, adapted from Ulrich et al., 2015). The automatic process has a characteristic time-to-peak which determines the influence of said process on the superimposed accumulation.

accumulation of evidence by the wrong accumulator larger than a pre-response threshold but not enough to fully trigger the response, as observed in electromyographic studies and LRPs. The two-phase LCA model implemented by Servant et al. (2015) does in fact present this behavior, although it fails at capturing the different delta plots.

The one model that successfully reproduces the delta plots is the Diffusion Model for Conflict tasks (DCM; Ulrich et al., 2015). The DCM is a dual-route extension of the DDM, where the accumulation of evidence is the sum of a controlled process, with constant drift coefficient, and an automatic process, with a time-dependent drift coefficient (see Fig. 2.9). The total drift is given by

$$\mu(t) = \mu_c \pm Ae^{-t/T} \left[\frac{t\,e}{(a-1)T} \right]^{a-1} \left[\frac{a-1}{t} - \frac{1}{T} \right] \tag{2.1}$$

where μ_c is the controlled process drift coefficient, A, a and T are parameters that determine the shape of the automatic process, and the sign of the second term determines whether it is a congruent (positive) or incongruent (negative) condition. The main, unique feature from the DMC is that by modifying the parameter T, which measures the time-to-peak of the automatic process, one can obtain different delta plot slopes. Fast automatic processes lead to negative slopes, while slow automatic processes lead to positive slopes. A parameter-recovery study by White et al. (2017) compares the DCM to other drift-diffusion models and concludes it is the one that can overall capture data from more tasks.

As Ulrich et al. (2015) admit, (2.1) is an entirely *ad hoc* equation with no neural basis and a very limited intuitive interpretation. However, because it is the only model flexible enough to reproduce the different slopes of delta plots from different Simon-like experiments, it is worth wondering whether similar results can be obtained with a more plausible, neurally-inspired model. In the next Section I show the answer is yes.

2.4. An application of MLCA to the Simon effect

From the overview of the experimental and computational results given in the previous Sections, one can conclude that the Simon effect is the result of a complex cognitive decision-making process. This process includes the generation of spatial codes for stimuli and responses, which involve attention and working memory, and some form of intentional control which inhibits an automatic response associated with the spatial codes. A successful neurocomputational account of the Simon effect must capture the differences in response times and error, produce partial errors for incongruent trials, result in delta plots with both positive or negative slopes, and provide an explanation of how the spatial codes interact.

In this section, I first present a summary of two initial unsuccessful attempts. Although these did not provide the expected results, they were useful to understand which elements are imperative to fulfill the previous requirements. Some details about these two models are spared and later explained when describing the triumphant dual-route model.

2.4.1. A tale of two failures

Brain-State-in-a-Box

Prior to experimenting with sequential sampling models, I tested a model inspired on the Brain-State-in-a-Box (BSB; Anderson et al., 1977). The BSB is a neural network based on recurrent autoassociative memories. Unlike heteroassociative memories described in Section 1.2.2, autoassociative memories store associations between an input vector and itself. In doing so, these models are particularly well suited for filtering or recovering an original vector from a noisy input vector. As an example, Kohonen (1977) uses autoassociative memories to store and later recover pictures of faces. When the system is presented with a distorted face, or a picture with missing pixels, it is able to produce the original one as an output. The BSB consists of a dynamical system, where the output is used as an input to the system in the following iteration. This is possible because in autoassociative memories, unlike heteroassociative ones, inputs and outputs necessarily have the same dimensions. The recurrent process continues until the output coincides with one of the corners of an m-dimensional hypercube of side 1 (see Fig. 2.10), which Anderson et al. (1977) call a box, where m is the number of orthogonal stored vectors.

The strength of the BSB comes from its robustness to recall information, even if the input differs from the stored vectors (see Fig. 2.10). My first model of the Simon effect was
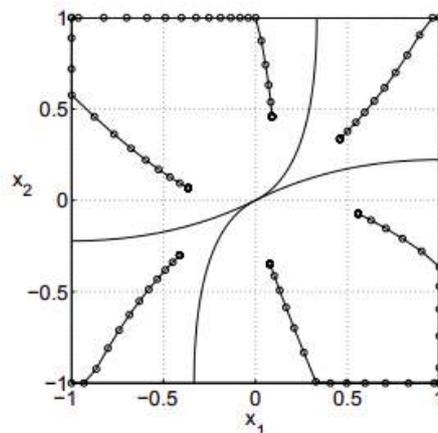
Figure 2.10: Dynamics of a BSB model in state space (Oh et al., 2005). The area of the unit square is divided into sectors such that any initial state within a sector converges to the corresponding corner.

inspired in a modified version of the BSB by Oh et al. (2005), but using context-dependent autoassociative memories (Pomi and Mizraji, 1999). Similarly to what was presented in Section 1.4, I suggest stimuli in a Simon task have two independent attributes: a spatial location $\boldsymbol{s}$ and a non-spatial feature $\boldsymbol{f}$. For example, for a Simon task like the one in Fig. 2.4, there are two feature vectors $\boldsymbol{f}_{square}$ and $\boldsymbol{f}_{circle}$, and two location vectors, $\boldsymbol{s}_{left}$ and $\boldsymbol{s}_{right}$. Unlike the experiment of Mante et al. (2013), the location of a stimulus is never relevant for the task; however, as explained previously, the spatial code is formed and somehow intervenes in the response selection process. The complete stimulus vector is therefore the Kronecker product of the feature and the location, $\boldsymbol{f} \otimes \boldsymbol{s}$. And, regardless of the location, that stimulus should be mapped into the response corresponding to the feature $\boldsymbol{f}$. I therefore considered an autoassociative matrix $\boldsymbol{M}$ given by

$$\boldsymbol{M} = \sum_i \eta_i \boldsymbol{f_i}(\boldsymbol{f_i} \otimes \boldsymbol{s_i})^T \tag{2.2}$$

where η_i is a scaling rate and the sum ranges over all possible associations. If the scaling rates for congruent conditions are larger than those of incongruent conditions (see Fig. 2.4) then the output when the matrix $\boldsymbol{M}$ is presented with a congruent stimulus is larger than with an incongruent one.

The output of the matrix is an updated feature vector $\boldsymbol{f}^{(t)}$, which is used as the input in the following interaction. The update rule for the feature vector is given by

$$\boldsymbol{f}^{(t+1)} = \boldsymbol{f}^{(t)} + \boldsymbol{M}(\boldsymbol{f}^{(t)} \otimes \boldsymbol{s}) \tag{2.3}$$

and integration continues until one of the components of $\boldsymbol{f}^{(t)}$ is equal to 1, i.e. until the vector reaches one of the sides of the box, and the corresponding response is selected. In this model, the feature vector is updated on each step of integration but the location vector is maintained constant, because the decision is made on the basis of the feature of the stimulus. Nonetheless, the location vector does affect the dynamics of this model through the output of the autoassociative matrix.

The preliminary results obtained with our BSB were positive. Response times were larger, i.e. slower, for incongruent trials than for congruent trials. Plus, if we included noise into the integration equation (2.3), then the error rates were also larger for incongruent trials. We also found that artificially increasing the difficulty of the task by decreasing the angle between the stored feature vectors resulted in a reduction of the Simon effect, as reported by Lammertyn et al. (2007). However, upon closer analysis, this model is unable to reproduce several key experimental results from the Simon effect literature. The main one is that the change in $\boldsymbol{f}^{(t)}$, which is given by the output of the matrix, is monotonous: for incongruent trials, the feature vector is deviated from the correct vector towards the wrong one. And because this model does not include any form of leakage or inhibition, this process continues until a response is made, meaning this model cannot explain partial errors. The model also fails to obtain negative slopes in delta plots. In a sense, this model is in a similar level as the one proposed by Zorzi and Umiltá (1995). It is a simple model, based on the intuitive idea that features and location are coded independently, but it can only explain some superficial experimental results.

Single-route MLCA model

After recognizing that the failure of the BSB model was due to the rudimentary dynamics of (2.3), I looked into other decision-making models. It was then that I decided to implement the LCA model of Usher and McClelland (2001) and developed the MLCA model introduced in Section 1.3. Following the successful reproduction of the experiment of Mante et al. (2013), I adapted the same model to try to reproduce the Simon effect. A

first attempt consisted in simply transcribing the equations from Section 1.4 using the same vectors defined for the BSB model, so that the context-dependent associative memory becomes

$$
\begin{aligned}
\boldsymbol{M} = \ &\eta \ \boldsymbol{r}_{left} \ (\boldsymbol{f}_{square}{}^T \otimes \boldsymbol{s}_{left}{}^T)+ \\
&\eta' \ \boldsymbol{r}_{left} \ (\boldsymbol{f}_{square}{}^T \otimes \boldsymbol{s}_{right}{}^T)+ \\
&\eta' \ \boldsymbol{r}_{right} \ (\boldsymbol{f}_{circle}{}^T \otimes \boldsymbol{s}_{left}{}^T)+ \\
&\eta \ \boldsymbol{r}_{right} \ (\boldsymbol{f}_{circle}{}^T \otimes \boldsymbol{s}_{right}{}^T)
\end{aligned}
\tag{2.4}
$$

where $\eta > \eta'$, and the accumulation process is given by

$$
\boldsymbol{dx} = [\boldsymbol{M}\,(\boldsymbol{f} \otimes \boldsymbol{s}) - \boldsymbol{Dx}]\,\frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}}
\tag{2.5}
$$

This first sequential sampling model had a performance similar to that of the BSB, in that it successfully discriminated between congruent and incongruent conditions. I tried to go further with this single-route model by incorporating a time-dependence to the stimuli, drawing inspiration from the sequential sampling model of Servant et al. (2015). I implemented a two-phase spatial location vector, with a time parameter T indicating the end of phase one such that, for a stimulus presented in the left hemispace, $\boldsymbol{s} = \boldsymbol{s}_{left}$ if $t < T$ and $\boldsymbol{s} = \boldsymbol{s}_{center}$ if $t > T$, where $\boldsymbol{s}_{center}$ is a normalized vector equidistant from $\boldsymbol{s}_{left}$ and $\boldsymbol{s}_{right}$.

By adding this time-dependence, the model becomes able to capture one of the most challenging experimental results of the Simon effect: partial errors. For incongruent trials, the wrong accumulator has higher evidence input during the first phase; this is reverted during the second phase, but if enough evidence is accumulated at that stage, a partial error akin to those detected by electromyography or LRPs occurs.

As anticipated from the work of Servant et al. (2015), a single-route sequential sampling model is unable to achieve negative slopes in delta plots. For this reason, the single-route MLCA model had to be discarded and replaced by a dual-route one. Nevertheless, this failed attempt led me into what ultimately became a key finding: because of leakage and inhibition, a two-phase MLCA model with high input during the first stage and low input during the second one has a qualitatively equivalent evolution to the automatic process used by Ulrich et al. (2015) in their dual-route model (see Fig. 2.9).

2.4.2. A successful dual-route MLCA model

There are two views on how a dual-route model may explain SRC effects. One possibility, detailed by Kornblum et al. (1990) and De Jong et al. (1994), and earlier intuitively implemented in the algorithmic model of Rosenbloom and Newell (1988), is that an automatic route processes the spatial information and prepares the motor program while a controlled route either confirms that response, in which case it is executed, or rejects it and prepares and executes the other response. Alternatively, the two routes can be independent and the automatic route is inhibited regardless of its output, whereas the slower, controlled route is effectively responsible for selecting a response. In order for this second account to explain the time difference between congruent and incongruent conditions, the two routes must be independent but their outputs must be superimposed at a central decision-making hub (see Fig. 2.9).

The available experimental data, such as the existence of partial errors or the asymmetric motor activity in LRPs, can be interpreted within both alternative possibilities. However, I decided to follow the second account for two reasons. First, the dual-route model of Ulrich et al. (2015), which is to date the computational model that

best fits the data, is based on two independent and superimposed processes. Second, I believe that the distributed accumulation of information represents the neural activity of the brain better than the algorithmic, symbolic processing. Ultimately, the point must be made that the very existence of the two routes is debated, so that only more and better techniques for measuring brain activity will provide the necessary proof. Here, I show one possible approach for reproducing the currently available experimental results.

In the following, I gradually unravel the dual-route MLCA model of the Simon effect, starting with the controlled process, then the automatic process, and finally the superimposed process. The overall goal is to reproduce the results of Ulrich et al. (2015) while maintaining biological plausibility.

Controlled process

The controlled process consists of the slow and monotonous accumulation of evidence towards the correct response. Only the stimulus' relevant feature determines the input to the accumulators – although due to the stochasticity of the integration, the incorrect accumulator may end up winning the race.

Using Fig. 2.4 as an example, I define two feature vectors, $\boldsymbol{f}_{square}$ and $\boldsymbol{f}_{circle}$. These two vectors need not be orthogonal. In fact, one can control the difficulty of the task by modifying the cosine similarity between them. A square and a circle are more dissimilar than a pentagon and a hexagon, so the angle between them must be larger. I also define two spatial location vectors, $\boldsymbol{s}_{left}$ and $\boldsymbol{s}_{right}$. Again, these location vectors need not be orthogonal, and their cosine similarity can be used as a measure of the physical distance between the two positions where the stimuli may appear. Finally, I define two response vectors, $\boldsymbol{r}_{left}$ and $\boldsymbol{r}_{right}$, which are orthogonal. All these vectors are assumed to be normalized.

A context-dependent associative memory for the controlled process, $\boldsymbol{M}_c$, is defined such that only the stimulus' feature determines the output:

$$
\begin{aligned}
\boldsymbol{M}_c = \; & \eta_c \; \boldsymbol{r}_{left} \; (\boldsymbol{f}_{square}{}^T \otimes \boldsymbol{s}_{left}{}^T) + \\
& \eta_c \; \boldsymbol{r}_{left} \; (\boldsymbol{f}_{square}{}^T \otimes \boldsymbol{s}_{right}{}^T) + \\
& \eta_c \; \boldsymbol{r}_{right} \; (\boldsymbol{f}_{circle}{}^T \otimes \boldsymbol{s}_{left}{}^T) + \\
& \eta_c \; \boldsymbol{r}_{right} \; (\boldsymbol{f}_{circle}{}^T \otimes \boldsymbol{s}_{right}{}^T)
\end{aligned}
\tag{2.6}
$$

where η_c is a scaling rate. Since the same scaling rate is used for all stored associations, the spatial information does not contribute to the controlled process. One could also use a linear associative memory that directly stores associations between features and responses; however, to maintain a coherence with the automatic process, I decided to use both attributes as inputs.

During an experimental trial, the presentation of a stimulus $(\boldsymbol{f}, \boldsymbol{s})$ triggers the accumulation for the controlled process according to the MLCA model, given by

$$
\boldsymbol{dx}_c = [\boldsymbol{M}_c \, (\boldsymbol{f} \otimes \boldsymbol{s}) - \boldsymbol{D}_c \boldsymbol{x}_c] \, \frac{dt}{\tau} + \boldsymbol{\xi}_c \sqrt{\frac{dt}{\tau}}
\tag{2.7}
$$

where $\boldsymbol{x}_c$ is an accumulation vector for the controlled route, $\boldsymbol{D}_c$ is the decay matrix and $\boldsymbol{\xi}_c$ is the Gaussian noise. The accumulation of evidence only finishes when the superimposed process reaches a decision threshold.

Automatic process

The second component of a dual-route model of the Simon effect is the automatic route. This is likely the most important element, for it is the automatic route that determines

the slope of the delta plots. However, since the dual-route model is a theoretical construct, there is no experimental evidence to determine its characteristics. Instead, one can try to reproduce the DMC model of Ulrich et al. (2015). Because the DMC captures the behavioral results of the Simon effect and because the model can produce both positive and negative delta plot slopes, the automatic route of Ulrich et al. (2015) is deemed a satisfactory one. I cannot deny its success, but must also be critical of its excessive mathematical specificity. Not only does the drift coefficient for the automatic process in (2.1) include several parameters and both polynomial and exponential time dependencies, the authors also fail to explain how the brain might implement such a drift function. The DMC is an example of overfitting due to excessive abstraction from the neural substrate, a constant risk in the field of mathematical psychology. Nonetheless, the work of Ulrich et al. (2015) can serve as a guiding light. Here, I intend to achieve similar behavioral results with a more biologically plausible neurocomputational mechanism.

The automatic process of the DMC has three parameters which control the three characteristics of the central curve in Fig. 2.9. These are the amplitude, the time-to-peak, and the decay. The time-to-peak is particularly relevant because fast automatic processes lead to negative slopes in the delta plots, while slow automatic processes lead to positive slopes. As hinted when explaining my failed single-route model, one can include a time-to-peak as a change between two different phases. During a first phase, a positive input means evidence is being accumulated; at time T, the input becomes null so there is no more evidence to accumulate. Also, because I am using the LCA as a basis for the model, the decay of the automatic process can be attributed to leakage. Since the DDM does not include leakage, Ulrich et al. (2015) are forced to add decay to their automatic process. Finally, it is also possible to control the amplitude of the automatic route by means of the scaling rates which define the memory matrix.

In conclusion, using the same vectors defined for the controlled process, I define a context-dependent associative memory for the automatic process as

$$
\begin{aligned}
\boldsymbol{M}_a = \ &\eta_a \ \boldsymbol{r}_{left} \ (\boldsymbol{f}_{square}{}^T \otimes \boldsymbol{s}_{left}{}^T) + \\
&\eta_a' \ \boldsymbol{r}_{right} \ (\boldsymbol{f}_{square}{}^T \otimes \boldsymbol{s}_{right}{}^T) + \\
&\eta_a' \ \boldsymbol{r}_{left} \ (\boldsymbol{f}_{circle}{}^T \otimes \boldsymbol{s}_{left}{}^T) + \\
&\eta_a \ \boldsymbol{r}_{right} \ (\boldsymbol{f}_{circle}{}^T \otimes \boldsymbol{s}_{right}{}^T)
\end{aligned}
\tag{2.8}
$$

where η_a and η_a' are the scaling rates, with $\eta_a > \eta_a'$. There are two important differences between this matrix and the one used for the controlled process. First, I use two different scaling rates because, at least in principle, both the stimulus' feature and the location contribute to the automatic process. Second, the associations for the incongruent conditions are shifted. For the automatic process, spatial location indicates the preferred response.

Now, during an experimental trial, the presentation of a stimulus $(\boldsymbol{f}, \boldsymbol{s})$ triggers the accumulation for the automatic process according to the MLCA model, given by

$$
d\boldsymbol{x}_a = [\boldsymbol{M}_a \, (\boldsymbol{f}^* \otimes \boldsymbol{s}) - \boldsymbol{D}_a \boldsymbol{x}_a] \, \frac{dt}{\tau} + \boldsymbol{\xi}_a \sqrt{\frac{dt}{\tau}}
\tag{2.9}
$$

where $\boldsymbol{x}_a$ is an accumulation vector for the automatic route, $\boldsymbol{D}_a$ is the decay matrix, $\boldsymbol{\xi}_a$ is the Gaussian noise, and $\boldsymbol{f}^* = \boldsymbol{f}$ if $0 < t < T$ or $\boldsymbol{f}^* = \boldsymbol{0}$ otherwise. During the first phase, the accumulation process has a fixed input, just as in the controlled route. However, during the second phase, the input becomes zero and the accumulation process becomes a decay process. (2.9) can be rewritten as

$$\boldsymbol{dx_a} = \left[\boldsymbol{M_a}\left(\boldsymbol{f} \otimes \boldsymbol{s}\right) - \boldsymbol{D_a}\boldsymbol{x_a}\right]\frac{dt}{\tau} + \boldsymbol{\xi_a}\sqrt{\frac{dt}{\tau}}, \quad \text{if } 0 < t < T$$

$$\boldsymbol{dx_a} = - \boldsymbol{D_a}\boldsymbol{x_a}\frac{dt}{\tau} + \boldsymbol{\xi_a}\sqrt{\frac{dt}{\tau}}, \qquad\qquad \text{if } t > T \tag{2.10}$$

where the second phase is simply a stochastic decay, at a rate given by the matrix $\boldsymbol{D_a}$.

Superimposed process

Having described the controlled and the automatic routes independently, the superimposed process is very straightforward. At each step, the two increments are added together, so that

$$\boldsymbol{dx} = \boldsymbol{dx_c} + \boldsymbol{dx_a} \tag{2.11}$$

where $\boldsymbol{dx}$ is the accumulation in the superimposed route and $\boldsymbol{dx_c}$ and $\boldsymbol{dx_a}$ are obtained from (2.7) and (2.9). The entire accumulation process finishes once the superimposed process reaches a threshold value or once a maximum time has elapsed.

Next, I present some numerical simulations to verify whether this dual-route model reproduces the expected behavior. Details about the numerical simulations are included in Appendix B. A useful first step is to visualize the temporal evolution of each of the two routes, as well as the superimposed process. Fig. 2.11 shows an example of an incongruent trial with an early time-to-peak. The automatic process guides activation in favor of the incorrect response during the first phase, but this subsequently decays and the correct response is ultimately selected. It is relevant to notice how fluctuations due to the addition of Gaussian noise play an important role in the MLCA model, in that the incorrect response is not too far away from the threshold for the superimposed process. Also, the two routes in the model are independent and do not interact with each other. Therefore, the high, early accumulation of evidence favoring the incorrect response in the automatic process does not inhibit the time evolution of the correct response in the controlled process.

Fig. 2.12 shows the paths in state space for that same numerical simulation. Again, one can see that the automatic process initially leads the path of the superimposed process towards the top boundary, i.e. the incorrect response. In this example, both the controlled and the automatic process have nearly perfect discrimination. Neglecting the fluctuations due to the noisy integration, the controlled process has a horizontal path towards the correct response while the automatic process has a vertical path, initially towards the incorrect
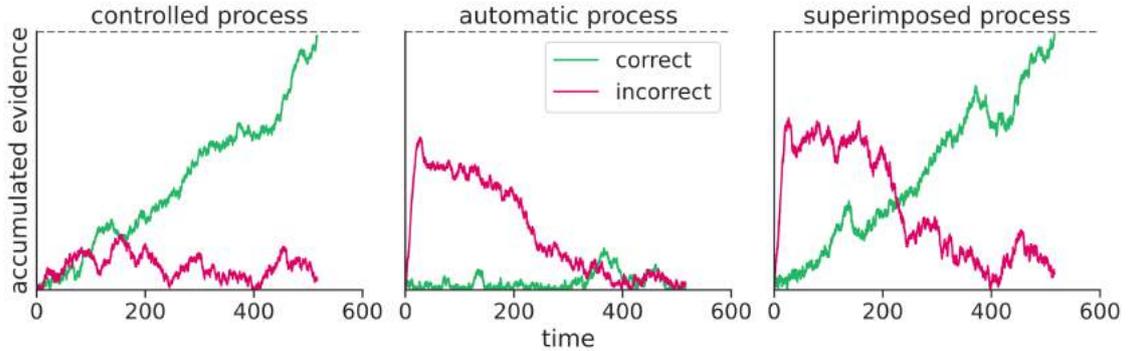
Figure 2.11: Example of dual-route MLCA simulation of the Simon effect. The automatic process has an early time-to-peak and then decays. A response is selected when the correct accumulator in the superimposed process reaches the decision boundary.
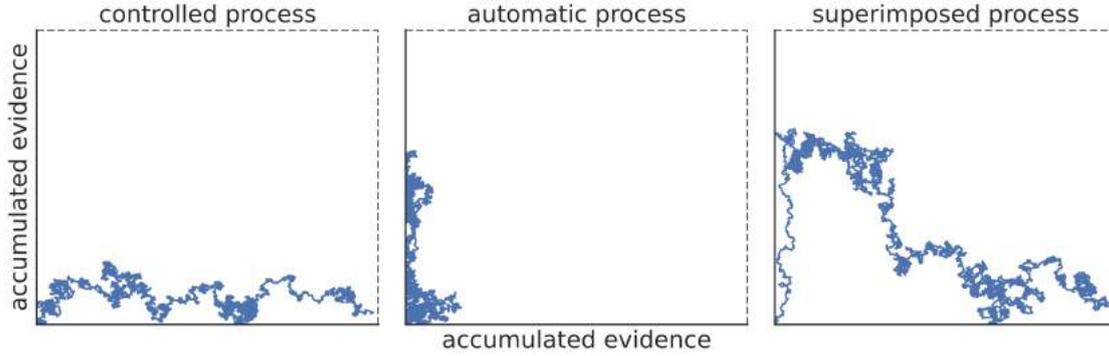
Figure 2.12: Example of dual-route MLCA simulation of the Simon effect in state space, with the same trial from Fig. 2.11. The horizontal axes indicate the accumulated evidence for the correct accumulator and the vertical axes for the incorrect accumulator.

response and then decaying. Importantly, the superimposed process, which is the only one that can be measured experimentally, combines these two behaviors and therefore presents a partial error.

Delta plots

Finally, it is worth looking at the delta plots obtained with the dual-route MLCA model. These are constructed as follows: first, numerical simulations are performed for both the congruent and the incongruent conditions. Next, the response times for each condition are split into bins. Finally, the response times from incongruent and congruent conditions are subtracted for each corresponding bin, and each of these values is plotted against the average of the two response times. Experimenters usually use five bins, or quintiles, because of limited amounts of data. Since this work consists of numerical simulations, we do not share that limitation.

The main finding of De Jong et al. (1994) is that the Simon effect, unlike other SRC effects, has delta plots with negative slopes. Salzer et al. (2017) find that the slope varies from one Simon task to another, and there are tasks for which the delta plots present positive slopes. Ulrich et al. (2015) successfully reproduce the different types of delta plots by changing the parameter T, which represents the time-to-peak in 2.1.

The dual-route MLCA model of the Simon effect can also reproduce both positive and negative slopes in delta plots. This is done by modifying the parameter T, which represents the end of the first phase for the automatic process, in (2.10). Fig. 2.13 shows a delta plot with a positive slope, obtained with a late time-to-peak of $T = 400$; Fig. 2.14 shows a delta plot with a negative slope, obtained with an early time-to-peak of $T = 20$. The stimuli features and locations for both figures are the same. Other than modifying the time-to-peak, it is also necessary to change the amplitude of the automatic process, so that expected accumulated evidence at the end of the first phase is equal for both early and late time-to-peak. This is controlled by the parameters η_a and η_a'. Alternatively, one could redefine these parameters to have an explicit dependence on T.

These results place the dual-route MLCA model alongside the DMC of Ulrich et al. (2015) as the only two models of the Simon effect capable of reproducing delta plots with both positive and negative slopes. Of course, these two models share several common elements since the dual-route MLCA is strongly inspired in the DMC. Both are sequential sampling models with two independent accumulation processes, one with a fixed accumulation rate and the other with a nonmonotonous time dependence. However,
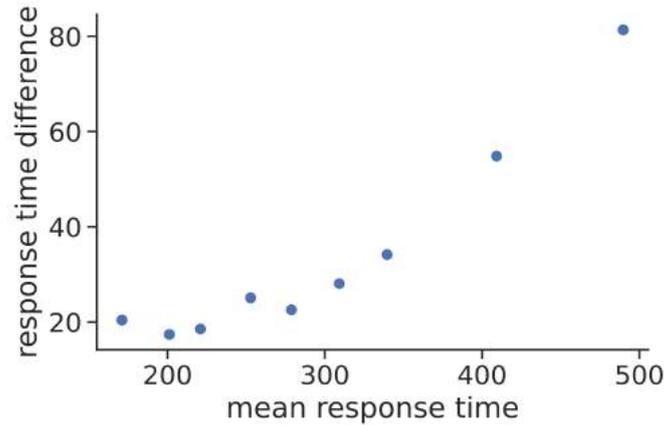
Figure 2.13: Delta plot with positive slope from MLCA simulations. These results indicate a higher Simon effect for slower responses due to a late time-to-peak in the automatic route.
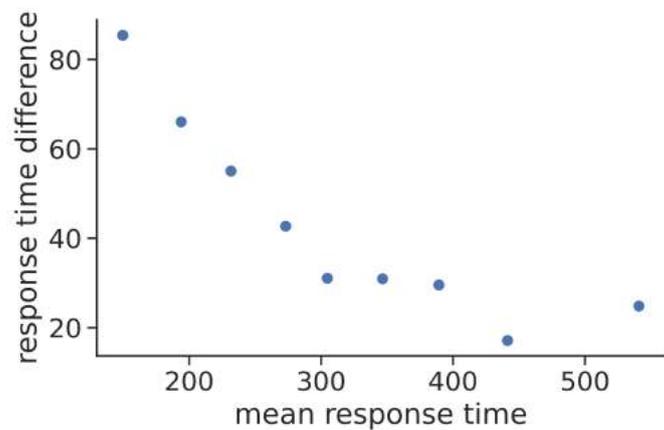
Figure 2.14: Delta plot with negative slope from MLCA simulations. These results indicate a higher Simon effect for slower responses due to an early time-to-peak in the automatic route.

there is one key difference between the DMC and the dual-route MLCA. The drift of the DMC has an explicit time dependence, given by (2.1), which can be fitted to experimental data in order to approximate response times nearly perfectly. The interpretation of those parameters is, however, not straightforward (see White et al., 2017). On the other hand, the time evolution of the automatic process in the dual-route MLCA is not forced but inherited from LCA dynamics.

2.5. Discussion

The spatial position of stimuli is encoded in several cortical and subcortical brain areas (Grieves and Jeffery, 2017), including the IPS in the PPC (Andersen et al., 1997; Azañón et al., 2010). The automatic activation of a spatial representation for stimuli can interfere with the execution of actions (Proctor and Vu, 2006), leading to slower responses when more recoding is necessary (Fitts and Seeger, 1953) across one or more overlapping dimensions (Kornblum et al., 1990), a phenomenon which receives the name of SRC.

One particular case of SRC, called the Simon effect, is due to incompatibilities between the hemispatial position of a stimulus with respect to the hemispatial position of the corresponding response (Simon and Small, 1969; Simon, 1969; Wallace, 1971). Unlike other SRC effects, the Simon effect has delta plots with positive and negative slopes depending on sensory modality and experimental settings (De Jong et al., 1994; Salzer et al., 2017). Several computational models can capture some experimental results (Zorzi and Umiltá, 1995; Zhang et al., 1999; Servant et al., 2015), but only dual-route models can account for negative slopes theoretically (De Jong et al., 1994) and computationally (Ulrich et al., 2015).

In this Chapter, I adapt the MLCA model to include two independent routes. One, called the controlled process, has a constant input into the accumulators; the other, called the automatic process, has a time-dependent input. A superimposed process, resulting from the addition of the accumulated evidences for both routes, is considered to determine the decision-making process.

Numerical simulations performed with the dual-route MLCA reveal that accumulation is initially guided by the automatic process but after some time, which depends on the time-to-peak parameter of the model, the automatic process decays and the controlled process becomes responsible for the response selected. A distributional analysis shows different types of results depending on the time-to-peak: if the automatic process is slow, i.e. long time-to-peak, then the delta plots have positive slopes; otherwise, if the automatic process is fast, i.e. short time-to-peak, then the delta plots have negative slopes. These results show that the MLCA model is capable of successfully reproducing the Simon effect.

Negative slopes in delta plots had previously only been obtained by a dual-route DDM model, the DCM (Ulrich et al., 2015). The DCM also has an explicit time dependence on the automatic route, although this is achieved by forcing a solution to the integration equation (2.1) rather than allowing the own dynamics of the model to determine the integration. The DCM has more flexible parameters which can be readily fitted to experimental data, although the interpretation of those parameters in relation to the mental and neural processes is unclear.

On the other hand, the automatic process in the dual-route MLCA model has a less intuitive time dependence. Rather than explicitly forcing the shape of the automatic route, I suggest an initial leaky accumulation followed by a decay is a good enough approximation to Fig. 2.9, particularly when adding Gaussian noise.

Two other possibilities were initially considered and must be mentioned at this point. First, instead of a two-phase automatic route, I tested an explicit time dependence on the spatial codes, in line with the hypothesis of decaying spatial codes (Hommel, 1994). This model had an accumulation equation equivalent to (2.9), where $\boldsymbol{f}^*(t) = e^{-t/T}\boldsymbol{f}$, where T was now the parameter that captured the time-to-peak. One could also increase the complexity of this explicit time dependence by considering an initial phase in which the spatial code is fixed, and a second phase in which it decays. A second alternative that I considered was to remove all time dependence from the automatic route and attribute its decay to the dynamics due to inhibition from the controlled route. Because the controlled process is slower, the automatic process would have higher activation during the initial

stages of integration. This model shares some conceptual commonalities with the original proposal for two processes (Kornblum et al., 1990). Ultimately, the suggested dual-route MLCA model and the two alternatives, one with spatial code decay and another with dual-route inhibition, all share the qualitatively equivalent behavior for the automatic route: an initial increase in activation followed by a decay. When one considers that Gaussian noise is added on every step of integration, the choice for one of the three models becomes arbitrary. Or, to be more precise, one chooses the alternative that is easiest to implement and interpret, knowing that the predictions are equivalent to those that would result from any of the other alternatives.

The success of the MLCA model to account for complex multi-attribute effects such as the Simon effect vouches for the interest in adapting it to also reproduce other known SRC effects as well. In line with the objectives of this thesis, the next Chapter is devoted to the SNARC effect, in which incongruencies between a number's magnitude and the position of its corresponding response lead to slower response times.

Chapter 3

On the spatial representation of numbers

In this Chapter, I apply the MLCA model to reproduce the main experimental result related to the number-space association: the SNARC effect. In Section 3.1, I argue for a cautious approach which requires a differentiation between purely numerical and spatial-numerical experiments. I then proceed to analyze these separately. In Section 3.2, I present the mental representation of numbers. In Section 3.3, I review experiments with an explicit spatial mapping of numbers. In Section 3.4, I review the implicit spatial mapping of numbers. Finally, in Section 3.5, I present a theoretical and computational account of spatial-numerical associations, including simulations for the numerical distance and the SNARC effects.

3.1. Three distinct number lines

A long list of psychological experiments, starting as early as the 1960s, show that numbers are mentally represented as continuous, analog quantities (Restle, 1970; Moyer and Landauer, 1967; Sekuler and Mierkiewicz, 1977; Shepard et al., 1975; Hinrichs et al., 1981). Two numbers with a small difference are more difficult to compare than two with a large difference, indicating that the codes for numerical quantities follow a one-dimensional, ordered structure. The currently accepted representation of numbers is named the mental number line (MNL): numbers are mentally represented as continuous, overlapping distributions on a one-dimensional, compressed scale (Dehaene, 2003). These mental representations originate from a neural substrate: number neurons with continuous, overlapping distributions of firing rates on a one-dimensional, compressed scale (Nieder and Miller, 2004).

Although the MNL refers to the representation of numbers in a mental, or neural, domain, it is not uncommon for researchers to attribute physical qualities to it. For example, when describing how infants perform approximate arithmetic, Opfer et al. write: "A mental number line makes basic addition and subtraction trivial: traveling four spaces forward from four registers the sum of four and four, traveling four spaces back from eight registers the difference between eight and four, and so on. Thus, if infants encode the approximate numerical value of a set and possess something akin to a mental number line (for at least nonsymbolic numbers), they should be able to register sums and differences of numeric quantities (at least approximately)." (Opfer et al., 2018, p. 110).

Is it possible to *travel four spaces forward* in an abstract mental representation? I believe there is a detrimental abuse of terminology with the metaphor of a number line in the field of numerical cognition. In the Introduction to this thesis, I mentioned a few

seminal publications which fueled the interest in studying the representations of numbers and space, including Francis Galton's report of subjects who visualize numerals in defined locations (Galton, 1880) and the SNARC effect (Dehaene et al., 1993). These and other related works, which are discussed in detail later in this Chapter, have led to the suggestion that numbers are *universally* mapped onto space (e.g. Dehaene et al., 2008). However, this need not be equivalent to attributing a spatial nature to the MNL.

Zorzi et al. (2002) suggest that the spatial representation of numbers originates in a functional isomorphism from the MNL to physical lines. From Umiltà et al. (2009, p. 564): "it is worth pointing out that Zorzi et al. (2002) did not claim that the number line is represented as a visual line. Functional isomorphism implies that the representation of the number line is based on the same spatial metric that underlies the representation of visual lines but it does not require any shared neural mechanism."

I propose a fairly simple solution based on the idea of Zorzi et al. (2002): the existence of three distinct types of number lines. One, the mental number line (MNL), refers to the abstract representation of numbers in the mind and brain. Another is a physical or geometric number line, which I call a spatial number line (SNL), where numbers are explicitly mapped onto space with well-defined spatial properties, such as shape, size, position, and direction. The last one, which I call an *as-if* spatial number line (as-if SNL), is an abstract construct where a non-existing physical line is covertly formed between two anchor points and numbers are implicitly mapped onto it. What the SNARC and other related effects reveal is that the MNL appears to be mapped to SNLs and as-if SNLs alike.

In these terms, it is possible to formulate the conundrum of number-space mapping as follows: what is the nature of the functional isomorphism from the MNL to SNLs and as-if SNLs? Is it innate, acquired through experience, or governed by working memory? To answer this question, I first analyze the MNL, SNLs, and as-if SNLs separately, and then consider the evidence for the functional isomorphism, which I reproduce by means of the MLCA model.

3.2. The mental number line

The idea of a MNL was first proposed by Frank Restle in 1970. Previously, Moyer and Landauer had performed a number comparison task, where subjects had to select the largest of two numbers. They detected that the decisions were faster when the numerical difference of the numbers was large, and suggested that "the displayed numerals are converted to analog magnitudes, and a comparison is then made between these magnitudes in much the same way that comparisons are made between physical stimuli such as loudness or length of time" (Moyer and Landauer, 1967, p. 1520). Restle built on this idea and went further, by stating: "A fairly simple model that agrees with the data of this experiment supposes that the subject uses a number line, an analog system having distinctive markers" (Restle, 1970, p. 277).

The tasks used to study the MNL include number recognitions, number comparisons, and basic arithmetic operations. These tasks do not involve any explicit spatial mapping of numbers onto space. Therefore, the number line suggested by Restle (1970) does not have any inherent physical attributes. It is not a SNL, but rather a mental or neural way in which numerical quantities are represented as analog magnitudes.

3.2.1. An approximate sense of number

Before discussing the MNL, an important distinction must be drawn between the representations of symbolic and non-symbolic numerical quantities. When we perform complicated calculations, we are capable of obtaining exact results in much the same way a digital calculator would. How does this exact processing of symbolic numbers relate to the analog system described by Restle (1970)?

The triple-code model of numerical processing (Dehaene, 1992; Dehaene and Cohen, 1995) proposes the existence of three independent but interacting representations for numbers: an analog non-symbolic magnitude, visual arabic digits, and verbal number words (Fig. 3.1). Each of these representations serves specific purposes. For example, the verbal system is employed for the rote-memory learning of number word associations, e.g. multiplication tables, while the visual system is employed for numeral-related tasks, e.g. performing multi-digit divisions. The analog non-symbolic magnitude representation is responsible for fast number comparisons and approximate operations, such as those measured by Moyer and Landauer (1967) and Restle (1970).

Unlike the other two systems, the analog magnitude representation is believed to be innate and universal. It is also available from birth for both humans and non-human animals. Two different core knowledge systems (Feigenson et al., 2004) are proposed to form the analog magnitude representation: the object tracking system (OTS), with which up to three or four objects can be counted and located immediately; and the approximate number system (ANS), with which approximate, non-symbolic quantities can be estimated and compared. Hereinafter, I refer to non-symbolic, approximate magnitudes as *numerosities* and restrict the word *number* to indicate symbolic, exact quantities.

One popular developmental hypothesis states that number symbols and words are mapped onto the OTS and the ANS, so that their meaning emerges from the magnitude representations (Spelke and Kinzler, 2007; Feigenson et al., 2004). This hypothesis is consistent with the speed and accuracy differences from symbolic comparison tasks (Moyer and Landauer, 1967; Dehaene et al., 1990). Whether or not symbolic abilities are based on magnitude representations is under debate (Wilkey and Ansari, 2019; Le Corre and Carey, 2007; Sasanguie et al., 2017). I refrain from commenting about said debate. Instead, I focus exclusively on the magnitude representation system, and particularly on the ANS, as a candidate for the MNL.

From introspective analysis, we can all identify an ability to produce gross estimates
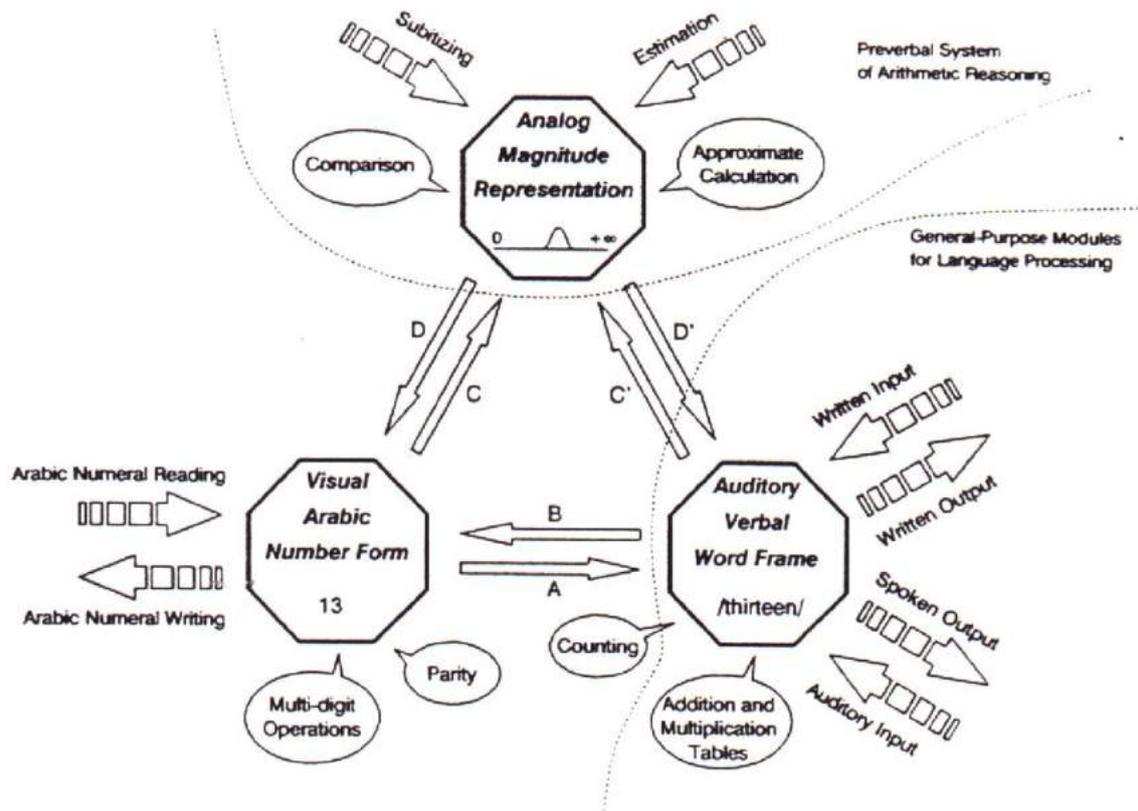
Figure 3.1: The-triple code model of numerical-processing (Dehaene, 1992). Three independent but related systems deal with different aspects of numbers. In particular, parity is coded by the visual arabic number form whereas magnitude is coded in an analog representation.

of people in a crowd or jelly beans in a jar within seconds and without counting. These results will not be exact, but more likely than not will be of the correct order of magnitude. For example, the title of this thesis has 93 characters, including spaces. One could have quickly estimated somewhere between 80 and 120 characters, but not 20 nor 500. This ability is the core knowledge called ANS.

The ANS has been reported in non-human animals, including monkeys (Cantlon and Brannon, 2007) and chicks (Vallortigara, 2017), as well as human adults (Halberda et al., 2012), newborns (Izard et al., 2009), and people with limited sensory input, such as congenitally deaf individuals (Spaepen et al., 2011).

3.2.2.   The numerical distance effect

The ANS, like other magnitude detection systems, is subject to Weber's law. Two different numerosities, for example two clouds of dots, can be accurately discriminated only if the ratio between the numerosities is larger than an individual-specific value, named Weber's fraction (Halberda et al., 2008). Weber's law is consistent with overlapping representations of the actual magnitude, which can be achieved on a linear scale with increasingly noisy representations (Gallistel and Gelman, 1992) or on a logarithmically compressed scale with fixed noise (Fechner, 1860; Dehaene, 2003). Although a linear scale is *a priori* a satisfactory solution, the logarithmic compression is regarded as more consistent with other mental and neural representations.

The overlapping mental representations were first proposed by Moyer and Landauer (1967) using a number comparison task and analyzing reaction times and percentage of
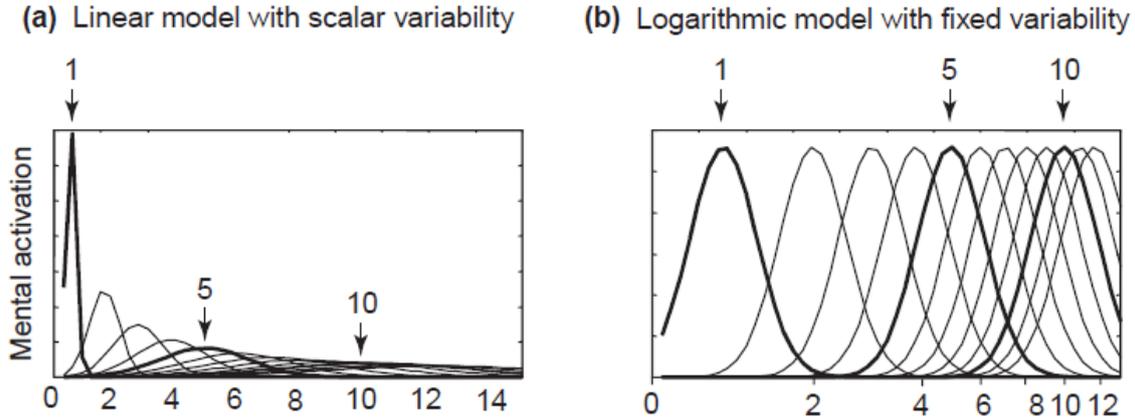
Figure 3.2: Linear and logarithmic models of the MNL (Feigenson et al., 2004). In both cases the total activation for each number is considered equal, which is why the absolute heights of the distributions in the linear case decrease.

errors as a function of the absolute difference between the two numbers. The same experimental paradigm was used by Sekuler and Mierkiewicz (1977) and Shepard et al. (1975). All these studies report faster response times and fewer errors as the difference between the two numbers increased. Restle (1970) combines a basic arithmetic task with a number comparison task, so that subjects have to choose the largest between a sum $A + B$ and a third number C. He analyzes response times and error rates as functions of different variables, and finds the most interesting result when plotting against the difference ratio $|A + B - C|/C$: a monotonously decreasing function. These results have been replicated numerous times (e.g. Hinrichs et al., 1981; Dehaene et al., 1990; Van Opstal et al., 2008).

The effect reported in these experiments is called the numerical distance effect, sometimes also called the comparison distance effect. In its simplest form, the numerical distance effect states that responses are slower and more error-prone as the numerical distance between the two numbers decreases. However, one must be cautious. The previous statement is true in general only if measured on a compressed scale. Alternatively, one can expand the original statement of the numerical distance effect to force that one of the two numbers remain fixed. The numerical distance effect is a consequence of Weber's law, as is another effect called the numerical size or magnitude effect: given a difference between the two numbers to be compared, responses are slower and more error-prone as the ratio between the two numbers decreases.

Upon reproducing Moyer and Landauer's results, Restle (1970) identifies that the numerical distance effect can be attributed to a one-dimensional representation, i.e. the MNL. On this MNL, each numerical magnitude has an analog representation with fixed noise, centered around the logarithm of said magnitude. Numbers on the MNL are ordered, so that the numerical distance effect is produced, and the overlapping representations are the cause of the slower, more error-prone, responses.

3.2.3. A number line without number words

Because of the mutual interactions between the three systems that form the triple-code model, accessing exclusively the MNL in literate adults can be difficult. However, because the ANS and the OTS are core systems that are available from birth in both humans and non-human animals, it is of great interest to study the MNL without number words or number symbols.

Cross-cultural differences

One way to access number representations without number words is by studying populations with primitive languages. For example, the Mundurucú are an indigenous group from an under-threat region between the states of Amazonas, Mato Grosso and Pará, in Brazil. They are partially educated, meaning some of their families receive or have received formal education in Portuguese, while others do not. Importantly, the Mundurucú have exact number words up to four, and beyond that threshold they use inexact expressions such as "one hand", "some, not many", or "many, really many". This is not rare among indigenous populations with no writing systems. Another Amazonian tribe, the Pirahá, only use the words for one, two and many (Gordon, 2004). A historical report about the Charrúa population in Uruguay indicates that they also only had a few number words (Gómez Haedo, 1937). Table 3.1 includes the number words used by these indigenous populations.

The results of Pica et al. (2004) indicate that, while Mundurucú participants achieve equivalent results to French controls in non-symbolic additions and comparisons with large quantities, they fail to perform approximate subtractions with small quantities beyond four. For this latter task, French participants are able to transform numerosities into symbolic numbers to perform operations; the Mundurucú, however, are limited by their lack of number words so they can only perform approximate arithmetic with numerosities.

The Mundurucú MNL, and therefore the innate human MNL, can only be employed for numerosity recognition, comparisons and arithmetic. Beyond the threshold of the OTS, the ANS cannot produce exact calculations.

Newborns, infants, and children

In a classical experiment, Wynn (1992) shows that 5-month-old infants are capable of performing exact arithmetic with very small numbers, e.g. 1+1 or 2-1, by employing
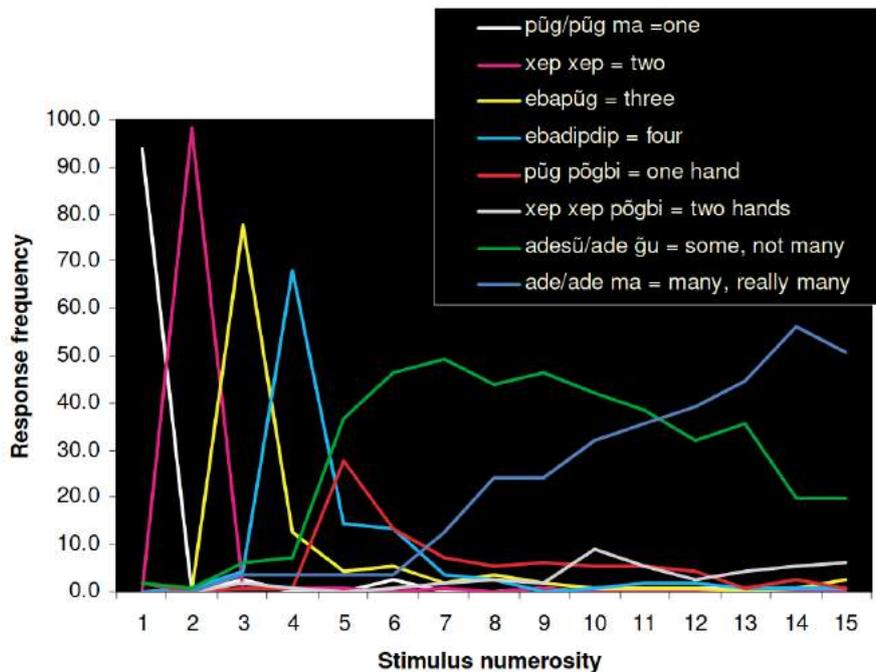
Figure 3.3: Number words in the Mundurucú language (Pica et al., 2004). They have exact number words up to four, but only words for approximate magnitudes for larger quantities.

Table 3.1: Number words from three indigenous South American languages: Mundurucú (Pica et al., 2004) and Pirahá (Gordon, 2004) from Brazil, and Charrúa (Gómez Haedo, 1937) from Uruguay.

Number	Mundurucú	Pirahá	Charrúa
1	pũgpũg ma	hói	Iú
2	xep xep	hoí	Sam
3	ebapũg	hoí	Detí
4	ebadipdip	hoí	Betum
5	adesũ/ade gũ	aibaagi	Betumiú
6+	adesũ/ade gũ	aibaagi	—

the OTS to follow dolls behind an occluder. As for the ANS, Xu and Spelke (2000) and McCrink and Wynn (2004) find that 6- and 9-month-old infants, respectively, can discriminate between large sets of stimuli if the ratio is large enough, e.g they discriminate between 8 and 16 but not 8 and 12, i.e. a ratio of 1:2.

Halberda et al. (2008) perform a numerosity discrimination task with children between 3 and 6 years of age, and find that the acuity in ANS changes during that time. 3-year-old children can discriminate up to a ratio of 3:4, whereas by 6 years of age they can discriminate ratios of 5:6.

Perhaps the most surprising finding with preverbal humans is the work done by Izard et al. (2009). They perform a numerosity discrimination from visual arrays and auditory tones with newborns of an average 49 hours of age. They find significant discriminations for 4 vs 12 and 6 vs 18, i.e. 1:3 ratio, but not for 4 vs 8.

Put together, the results from babies and children show that the ANS acuity develops from a starting 1:3 ratio to 1:2 around the first year of age and 5:6 by the beginning of formal schooling. Interestingly, the results from Pica et al. (2004) indicate that French adults have a Weber ratio of 9:10 and the Mundurucú have a slightly worse 6:7, which is comparable to that of 6-year-olds starting formal education. Higher ratios indicate a MNL with more spacing between numbers, or equivalently less noisy distribution, so it is possible to conclude that there is an ontogenetic development of the MNL from birth which is further enhanced by formal education.

Non-human animals

The idea of core knowledge systems being available universally and from birth requires, from an evolutionary perspective, that those same abilities must also be present at least in non-human primates (Spelke and Kinzler, 2007).

Cantlon and Brannon (2007) study the numerical competence of rhesus monkeys with a numerosity addition task. The monkeys are shown two sets of dots sequentially on a screen, and then have to select the result of the sum between two alternatives. Their performance depends on the ratio of the distractor, i.e. the incorrect response, with faster and more accurate responses when the alternatives are numerically different. This provides evidence for an analog magnitude representation in animals. Importantly, the same pattern of results is obtained when repeating the task in humans, although the overall performance of the latter is better, indicating a more advanced acuity of the ANS.

Another well-studied species with numerical abilities is the domestic chicken. By imprinting newborn chicks with objects, it is possible to study their discrimination

competence. Rugani et al. (2009) find that chicks can discriminate between two sets of up to 4 objects hidden behind screens, and select the group with the largest one. This discrimination falls within the limits of the OTS. Rugani et al. (2013) find similar results with large groups of stimuli, showing chicks can discriminate between 5 and 10 but not 6 and 9, which indicates an ANS acuity of 1:2.

Approximate number discrimination has been reported in dogs (Ward and Smuts, 2007), crows (Ujfalussy et al., 2013), and even fishes (Agrillo et al., 2012). Although it is unclear why or how numerical competence emerged, these results suggest that the evolutionary appearance of the ANS may predate the divergence of tetrapods from bony fishes (Agrillo et al., 2012).

3.2.4. The neural representation of numbers

Up to this point, I have discussed the behavioral evidence for a MNL in both humans and non-human animals. However, number-based tasks involve a myriad of functions, from perception to motor actions. If numerical quantities have their own mental representation in the form of a MNL, then a specific neural substrate must be employed. Here, I list some of the main neuroimaging and electrophysiology studies that reveal the neural basis of numerical cognition.

A fronto-parietal network

The first suggestions of possible brain regions involved in number representation came from patients with calculation deficits following brain damage, i.e. acalculia. Gerstmann (1940) identifies that lesions between the parietal and the occipital cortices lead to acalculia, among other conditions. Luria et al. (1966) identify the frontal lobe as being responsible for several cognitive functions, including arithmetic. Using positron emission tomography (PET), Dehaene et al. (1996) examine the brain areas activated during multiplication and number comparison. They identify a bilateral activation of parietal and prefrontal areas. Importantly, they find asymmetrical differences between multiplication and comparison, indicating task-specific representations. A meta-analysis of functional magnetic resonance imaging (fMRI) studies that use number comparison and calculation reveals a large network of parietal and prefrontal regions with task-specific differences (Arsalidou and Taylor, 2011; Fig. 3.4).

Several studies investigate how the neural representations of numbers depend on the way these are presented. According to the triple-code model, the word *four*, the numeral *4*, and the array of dots ⸬, each have their own neural patterns, although they should all activate a common brain region responsible for analog magnitudes. This is precisely what Piazza et al. (2007) look into during an fMRI study with French adult subjects. The study consists of an adaptation experiment, where subjects see a sequence of adaptation numbers (e.g. 17, 18, 19) for 2 minutes, and then the sequence of numbers changes abruptly (e.g. 47, 48, 49). During different trials, numbers are presented either as arabic digits or sets of dots. They find that, regardless of notation, the anterior regions of the IPS in the parietal cortex are activated whenever the numerical quantities change. They conclude that the human IPS is the best candidate for the neural coding of abstract numbers.

Number neurons

Single-cell recordings in monkeys and other non-human animals have been the main source of evidence for the neural substrate of numerical cognition since the turn of the century (Nieder et al., 2002; Nieder and Miller, 2004). The experimental paradigm employed by Nieder et al. (2002), called a delayed match-to-sample, consists of a visual display in which a first sample set of dots appears on a screen, and after a delay a new set of dots appear; if
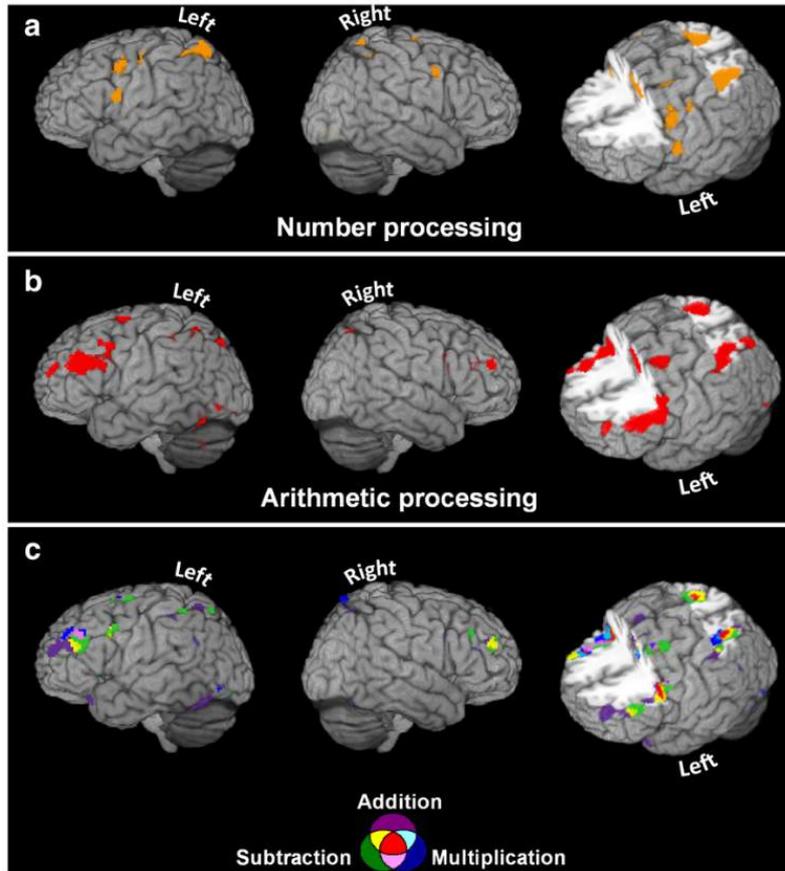
Figure 3.4: Human brain areas activated by different number tasks (Arsalidou and Taylor, 2011). The PPC and PFC are active for different types of tasks.

there is a numerosity match, then the monkey must respond, otherwise it must wait until a next set of dots. Monkeys trained on this task perform above chance with up to 5 items.

Nieder and Miller (2004) measure single-cell activity in several regions while monkeys perform the delayed match-to-sample (Fig. 3.5). They inspect neurons from the lateral prefrontal cortex (PFC), PPC, and anterior inferior temporal cortex. Within the PPC, they test different regions of the IPS. Each neuron is tested multiple times with items between 1 and 5, and their spike rates are averaged for each numerosity of items. They then analyze the time evolution of the spike rates for each neuron, in order to determine whether each specific neuron has a numerosity-dependent firing rate. They find around 30% of neurons in the lateral PFC and 20% of neurons in area VIP of the IPS are selective to numerosity. For other areas of the IPS and anterior inferior temporal cortex, fewer than 10% of neurons show any preference.

The latencies of the firing rates indicate that the IPS is activated before the PFC, so that numerical information may be first encoded in PPC and then conveyed to the PFC to guide the decision-making process. Importantly, numerosity-selective neurons in VIP and PFC represent abstract number, irrespective of modality. These neurons are responsive to numerical information presented simultaneously or sequentially (Nieder and Miller, 2004), as well as to visual and auditory stimuli (Nieder, 2012). Number neurons have a topographic organization similar to that of primary sensory cortices (Harvey et al., 2013).

Of particular relevance to the idea of the MNL is that number neurons are not only tuned to specific numerosities but also that the tuning is analog and follows Weber's law. Nieder (2016) argues that the tuning of number neurons is logarithmic, because only then
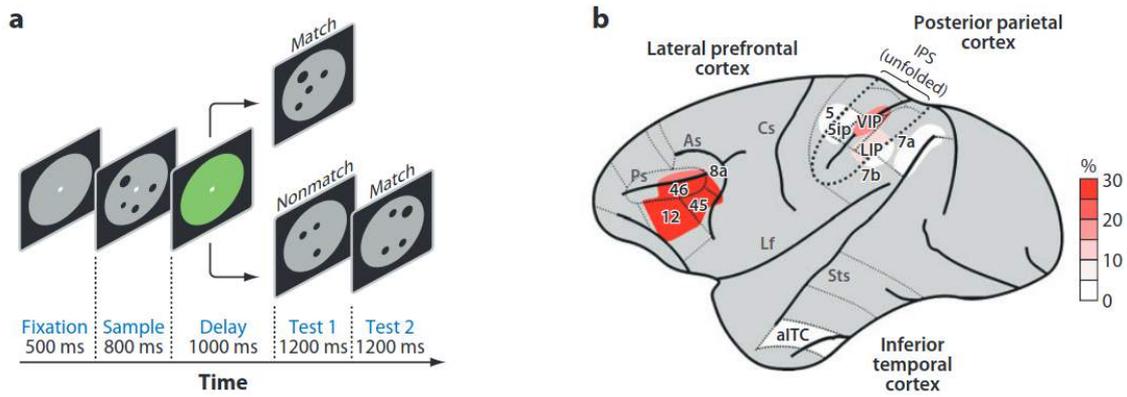
Figure 3.5: Detection of number neurons in the monkey brain (Nieder and Miller, 2004). Up to 30% of the neurons studied in the lateral PFC are tuned to specific numerosities.
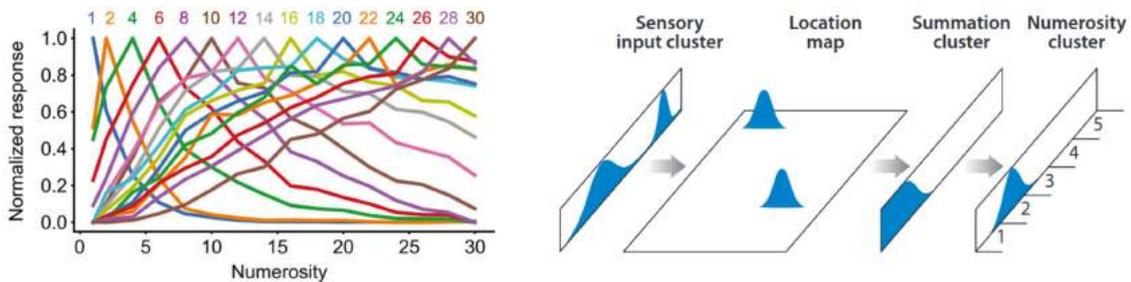
Figure 3.6: Number neurons *in silico*. Left: numerosity-selective neurons in a deep neural network (Nasr et al., 2019). The asymmetric distributions on a linear scale correspond to a logarithmic mapping (see 3.2). Right: summation units as an intermediate step for number-selective units (Dehaene and Changeux, 1993). Each of these hypothesized summation units integrate the inputs, for example from the visual field, and are active if *at least* a certain number of items is present.

do the tuning curves become symmetric Gaussian functions. The undeniable parallelism between number neuron tuning curves and behavior is taken as strong evidence in favor of a MNL grounded in area VIP of the IPS.

Number sense *in silico*

As a final note, I would like to mention two computational approaches that have contributed to the understanding and consolidation of the idea of a MNL.

The development of multi-layer artificial neural networks, also called deep neural networks, has had a powerful impact on different fields. Although there are obvious differences between these computational architectures and the animal brain, the training process of an artificial neural network has been compared to the statistical learning process of a newborn. Following this idea, Nasr et al. (2019) train a convolutional neural network to detect objects in images, and find that number-selective nodes emerge spontaneously. Remarkably, these number detectors share the characteristics of the number neurons found in monkeys, such as the analog selective tuning and logarithmic compression. The authors conclude that a number sense might be an emergent property of both artificial and natural visual systems.

An important point to make about the number neurons found in monkeys (Nieder

et al., 2002) and in artificial neural networks (Nasr et al., 2019) is that they are tuned as number-sensitive units, not as summation units. A summation neuron tuned to the number 7, for example, would have a high firing rate if there are *at least* 7 items. Dehaene and Changeux (1993) hypothesize that summation units can act as an intermediate stage between the detection of the item locations and the abstraction of their quantity (Fig. 3.6). That intermediate stage would be a simple way in which the brain could integrate both simultaneous and sequential number information prior to making a choice about the number. Verguts and Fias (2004) include summation units in a hidden layer of an artificial network trained to detect the number of active input units. Zorzi and Butterworth (1999) and Zorzi et al. (2005) use summation units to reproduce behavioral data from number comparison experiments. However, such summation units have not been detected experimentally. To the best of my knowledge, the identification of a numerosity of objects from visual inputs is performed without an intermediate summation.

3.3.   The spatial number line

A SNL is an explicit mapping of numbers on a line. We find SNLs in our everyday life. Rulers and thermometers are physical examples of straight SNLs. The axes of a Cartesian coordinate system in a two-dimensional graph are made up of two perpendicular SNLs. Clocks and speedometers are not straight lines, but are nonetheless one-dimensional, ordered arrangements of numbers, and therefore can be considered SNLs.

Because all possible ordered arrangements of numbers on a one-dimensional space are grouped into this category, I refer to each of these as a different SNL with its own properties. But there are some common subgroups. First, the linear SNL is one where the distance between two consecutive integers is constant throughout the entire line. This distance is called the unit length. On a linear SNL, the numerical difference between any two numbers is equal to the physical distance between them, on a scale given by the unit length. Similarly, the SNL on a clock face has a unit length corresponding to the arc length between two consecutive hours, and one can measure the time by calculating the arc length that separates the hour hand and the closest hour marker. Non-linear arrangements of numbers are also common. Logarithmic scales are used in graphs with information spanning over multiple orders of magnitude. Speedometers often show a compressed scale for large speeds. Finally, other arbitrary number lines do not follow any of these rules. The Ulam spiral is an arrangement of prime numbers on a square spiral, and can also be considered as a SNL.

The mappings of numbers onto a spatial line appear nowadays to be ubiquitous and universal (Dehaene et al., 2008). If so, then they must be innate, meaning they spontaneously emerge without the need for education. To determine whether SNLs are innate, it is necessary to review their evidence across the three different timescales of phylogenetic evolution, ontogenetic development, and cultural progress.

3.3.1.   A brief history of number-space mappings

Although non-human animals have some limited numerical abilities (see Sections 3.2.3 and 3.3.2), abstract mathematical knowledge is exclusive to humans. In fact, it is possible to trace back this development across civilizations by means of archeological findings (Fowler and Robson, 1998). The idea that number-to-space mappings are somehow natural forces one to look back at those early stages of mathematical knowledge. Surely, if numbers and space are universally linked, then the development of mathematical abstractions and reasoning must be supported by spatial representations.

One of the earliest remaining traces of sophisticated mathematical knowledge is the tablet YBC 7289 from the Babylonian civilization, approximately dating back to the year 1800 BCE (Fowler and Robson, 1998; Fig. 3.7). This clay tablet is engraved with a square with its two diagonals and the numbers 30, 1, 24, 51, 10, 42, 25, and 35, and it is believed to show an approximation of $\sqrt{2}$ in hexagesimal notation. YBC 7289 is one example of the level of understanding of abstract concepts and calculations that civilizations living four millennia ago were capable of. The shape drawn most likely reveals that Babylonians understood that $\sqrt{2}$ is the length of the diagonal of a unit square, and they used this geometrical link between number and space to both motivate their interest and improve their methodology to approximate its value. Other Babylonian tablets with sophisticated mathematical knowledge include Plimpton 322 and IM 67118 (Robson, 2002). However, among the more than half a million Babylonian tablets, 5000 of which contain mathematical knowledge, there are no number lines.

The Babylonians are one of the oldest civilizations from whom we have archeological remains including symbolic mathematical knowledge. But Robson (2002) claims most of the writing on their tablets refers to legal documents, such as contracts and transaction receipts, which might explain why there are no number lines. However, none of the other
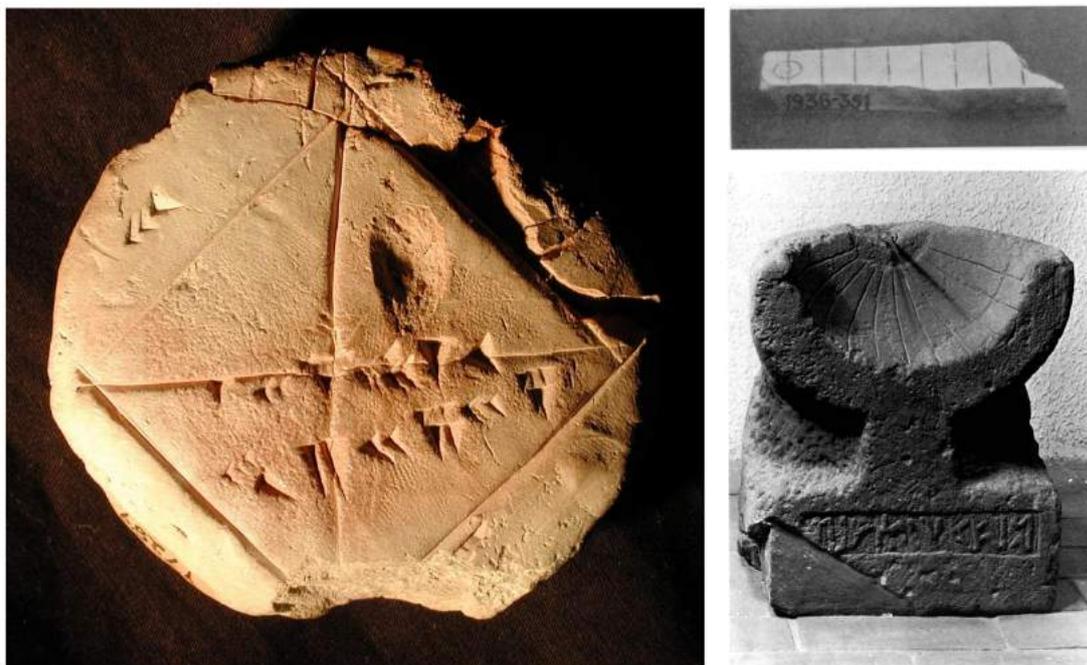
Figure 3.7: **Archeological findings revealing sophisticated geometrical knowledge of ancient civilizations**. Left: Tablet YBC 7289 from the Babylonian civilization (Fowler and Robson, 1998). Top right: Ivory ruler from the Indus Valley civilization (Sarton, 1937). Bottom right: Sundial from the Nabatean civilization (Healey, 1989).

advanced ancient civilizations with well-documented interests in theoretical mathematics and renowned engineering prowess left any written evidence of number lines, be it in Greece, China or the Middle East (Núñez, 2011). Although they had a profound understanding of numbers, they did not use SNLs to communicate or teach them.

Some of these ancient civilizations came extremely close to what one would nowadays call a SNL. For example, the Indus Valley civilization used ivory rulers between in 2400 and 1500 BCE (Sarton, 1937; Fig. 3.7). These rulers were found in Mohenjo-Daro, modern-day Pakistan, and had marked divisions every 3.35 cm although there were no numbers imprinted on them (Whitelaw, 2007). Another example is the sundial discovered in Mada'in Salih, modern-day Saudi Arabia, which was used by the Nabateans around the year 300 BCE (FIG. 3.7). Again, there are spatial lines indicating uniform divisions of time, but there are no numbers (Healey, 1989).

The first recorded instance of numbers being displayed as an ordered arrangement happened in the 11th century, with the invention of the clock face. Originally, clocks were water-powered and designed to make a sound every certain amount of time, usually by ringing a bell on top of a tower. The mathematician and physicist Ibn al-Haytham designed a mechanical system which included a rotating dial with marks for every hour (Wee, 2016). Unlike the sundial, the marks moved with the passage of time, so he was forced to identify them by adding numerals. It is unclear whether al-Haytham's clock should be considered a SNL because there are no reports of him intending to create a mapping of an ordered number sequence to a spatial configuration. By the 14th century, clock towers with clock faces were already popular across Europe (Lankford, 2013).

Interestingly, the first formal mention of a SNL used in mathematics is well-known (Núñez, 2011). In his 1685 book *A Treatise of Algebra*, the mathematician John Wallis introduced the concept of a number line to represent basic arithmetic operations. The

Figure 3.8: The first use of a number line in mathematics. Top: extract from Wallis (1685) where he first introduces the concept of a number line. Bottom: two diagrams from Descartes (1637) that reveal the highly abstract geometrical reasonings of the author, although no numbers are attributed to the letters and distances.

book includes a drawing of a SNL after the following paragraph: "Supposing a man to have advanced or moved forward (from A to B) 5 Yards; and then to retreat (from B to C) 2 Yards; If it be asked how much he had Advanced (upon the whole march) when at C? or how many Yards he is now Forwarder than when he was at A? I find (by Subducting 2 from 5) that he is Advanced 3 Yards (Because $+5 - 2 = +3$)" (Wallis, 1685; Fig. 3.8).

By 1685, many of the works which constitute the core of the 17th century scientific revolution had already been published, including Fermat's theory of numbers, Pascal's foundations of probabilities and pressure, and Napier's concept of logarithms. Leibniz published his development of differential and integral calculus in 1684. Both Kepler and Galileo made their discoveries about celestial bodies prior to SNLs. And, remarkably, Reneé Descartes' *La Géometrie* was published in 1637, nearly three decades before numbers were first mapped onto a spatial line. In fact, none of the 49 illustrations from said text include Cartesian coordinates. And some diagrams reveal how Descartes thought about geometrical relations between lines and shapes as being non-numerical ones (see Fig. 3.8).

The way in which John Wallis describes the idea of a number line within a book in which he also establishes the foundations of modern algebra and how to find the imaginary

roots of an algebraic equation is, from a modern day's perspective, astonishing. If Wallis' fellow mathematicians required such a detailed explanation, then certainly the concept of SNLs was not present within the general population. This is strong evidence against the idea of the innate or universal number-space mappings. From a historical perspective, SNLs have been manifest in human cultures for merely 350 years.

There is, however, a possible counterargument. Both in archeological findings from ancient civilizations and texts from a few centuries ago, geometric spatial knowledge is evident. Our diachronic analysis indicates that no SNL existed because there are no symbolic numerals explicitly associated with spatial lines. This may be because there was no interest or need for it. Next, I review experiments in which human and non-human animals are explicitly instructed to map numbers onto space.

3.3.2. Number line estimations in humans and animals

Number lines are a particularly powerful tool for the instruction of number comparisons, arithmetic, and a general sense of numbers during development (Woods et al., 2018). That is, young children trained on number lines show significant improvements in mathematical abilities (Lourenco et al., 2018; Hamdan and Gunderson, 2017; McCrink and Opfer, 2014). Here, I am not interested in how number lines influence development, but rather the opposite. The goal is to determine whether or not, and how, adults, children, and non-human animals are able to produce SNLs.

The number line estimation task consists of placing a target number somewhere in a line that has two anchor numbers, usually at the edges of said line. The result of this task is, by all accounts, a SNL. Although number line estimations have usually been used as manifestations of the MNL, in this Section I focus mainly on the spatial nature of the task.

In their classical study, Siegler and Opfer (2003) test children from second, fourth, and sixth grade, as well as adults, in a number line estimation task with numbers between 0 and 1000, which are the anchors at both ends of the line. Their results show that both children and adults have an inherent understanding of the ordinal mapping of numbers onto a spatial line, although there are differences between the groups. While adults and sixth graders present linear mappings, second and fourth graders spread small numbers and compress large numbers, which is best fit by a logarithmic mapping (Fig. 3.9). These results are in line with similar ones obtained through other numerical tasks. In line with Fechner's Law, our innate ability to estimate numerical magnitudes could be a logarithmic function of the magnitude, later transformed into a linear function by years of education (Dehaene, 2011). Siegler and Opfer (2003) repeat the experiment with second graders but changing the anchors to 0 and 100 and find a linear mapping (Fig. 3.9). By the second grade, i.e. around 7 years of age, the logarithmic-to-linear process is well underway, although only for relatively small numbers. Siegler and Opfer (2003) conclude that humans of different ages must have different numerical representations.

Those original results have been replicated numerous times. For example, Sella et al. (2020) perform a number line estimation task in first and second graders. In their task, however, a single anchor appears in the center of the line and the children have to place two target numbers, so that they are tested on their ordinal knowledge as well as their number-space mappings. The results indicate that the logarithmic-to-linear mapping for the number line correlates with other mathematical skills measured, such as arithmetic, but there is no such correlation for ordinal knowledge. The authors conclude, contrary to Siegler and Opfer (2003), that the number line estimation task is a measure of arithmetic ability and not of how numbers are represented.

Although the logarithmic-to-linear mapping is one of the many changes in visuospatial reasoning that children go through throughout education (Newcombe et al., 2018; Lourenco et al., 2018), the underlying logarithmic compression is still present in adults. Anobile
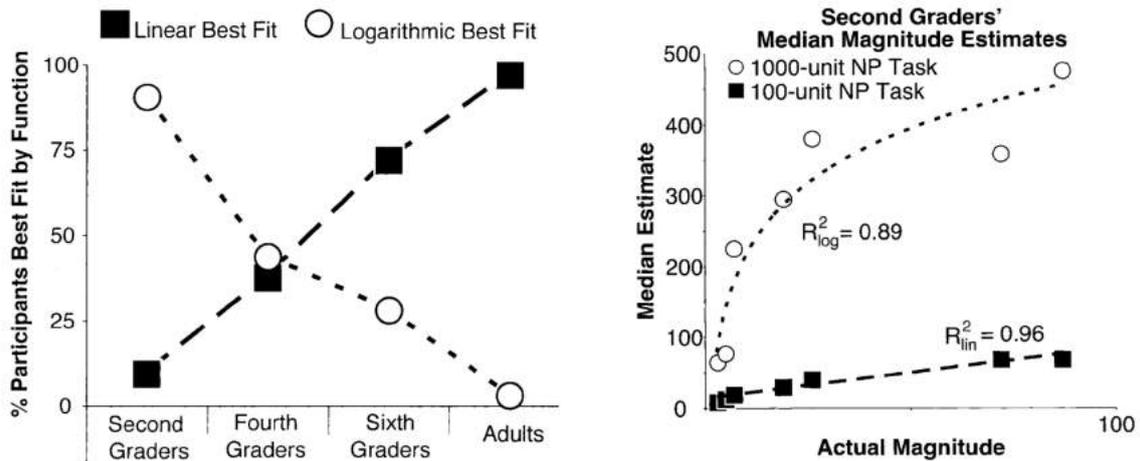
Figure 3.9: **Linear and logarithmic number line estimations across development**. (Siegler and Opfer, 2003). Left: as children grow, their behavioral results are better fit by a linear function than by a logarithmic one. Right: a linear fit is possible for second graders with smaller numbers.

et al. (2012) have adult subjects perform a number estimation task with clouds of dots. Under normal conditions, the mapping of numbers onto space is linear. However, when subjects perform an attentionally demanding task simultaneously, the mapping becomes logarithmic. This study indicates that linear SNLs may not be intuitive as they are strongly dependent on attentional resources.

Cross-cultural differences

The hypothesis of a logarithmic SNL as a more intuitive mapping for humans is supported by cross-cultural studies. Dehaene et al. (2008) test Mundurucú adults and children (see Section 3.2.3) on a number line estimation task with sets of dots, sequences of tones, spoken Mundurucú number words and spoken Portuguese number words. They also test American adults with dots, tones and English number words. Their results indicate that, while American controls alternate between linear and logarithmic mappings depending on the stimulus, Mundurucú participants show logarithmic mappings for all conditions.

Interestingly, Dehaene et al. (2008) report that some participants respond exclusively on the ends of the line presented. This constitutes a categorical rather than a spatial mapping. A similar result was obtained with the Yupno population in Papua New Guinea (Núñez et al., 2012; Cooperrider et al., 2017). Again, these results could be seen as contradictory to the idea of an intuitive spatial representation of numbers.

Number-space mappings in infants and animals

Unlike number comparison tasks, which are frequently used with infants and non-human animals (see Section 3.2.3), the number line estimation task is difficult to apply without language. Some similar experiments have been successfully implemented and reveal number-space mappings analogous to a SNL.

One common alternative to the number line estimation task is the line bisection task: participants are presented with a horizontal line and asked to indicate the center point, while the ends of the line have task-irrelevant flankers. De Hevia and Spelke (2009) use a line bisection task with symbolic numerals and non-symbolic clouds of dots as flankers. They test pre-school children with non-symbolic flankers, as well as adults with both types of flankers, and find a significant deviation of the bisection point towards the side ipsilateral

to the larger number. They conclude that, because young children with no formal education show the same pattern of results as adults, number-space mappings must be either innate or shaped during the first few years of experience.

De Hevia et al. (2014) study 7-month-old infants with a looking time habituation task. During the habituation stage, children are presented with non-symbolic numerical sequences from left to right that are either increasing or decreasing. During the test stage, they show significant differences in looking times between the familiar and novel sequence categories. De Hevia and Spelke (2010) perform a similar looking time habituation task with 8-month-old infants and find an association between number of dots and line length. Although these are not number line estimation tasks, they indicate that preverbal infants somehow associate numbers and space.

The closest experimental paradigm to a number line estimation with non-human animals comes from a series of studies with newborn chicks performed by Giorgio Vallortigara's group (Rugani et al., 2010; Rugani et al., 2007; Vallortigara, 2017). In their experiments, 4-day old chicks are placed on an apparatus with 10 aligned identical holes. During training, the chicks learn to peck at a specific hole, e.g. the fourth, to find food. During testing, the apparatus is changed to a different one with only 5 holes, and larger gaps between holes. Rugani et al. (2007) find that chicks peck significantly more at the fourth hole of the sequence, which corresponds numerically but not spatially to the hole they are trained with. The authors conclude that the use of ordinal information, e.g. counting four holes, is possible after just a few days of life. Similar results can be obtained with rhesus monkeys, which perform an adapted version of the experiment on a screen (Drucker and Brannon, 2014).

3.4. The *as-if* spatial number line

In the previous Sections, I described the MNL as an internal representation of numerical magnitudes and the SNL as an explicit mapping of the MNL onto space. These are two different number lines which occupy different representational domains, namely a mental or neural domain for the MNL and a physical domain for the SNL. In this Section, I review a series of experimental results which appear to question the strict dichotomous differentiation between MNL and SNL. Since the discovery of the SNARC effect (Dehaene et al., 1993), several researchers have suggested that the MNL itself has spatial attributes even when no explicit physical line is present. Alternatively, these results can be attributed to a mapping of the MNL onto space, as if a physical line was formed between the locations of the responses.

This idea is reminiscent of the somatic marker hypothesis (Damasio, 1996), whereby an emotion can be initially evoked by the body's response to a stimulus, called a *body loop*, and later the same emotion can be evoked without it, called an *as-if body loop*.

Similarly, I hypothesize a mapping of the MNL onto space as if there were a SNL. This leads to the necessary definition of a third category of number line: the *as-if* SNL. The as-if SNL is an implicit mapping, such that numbers are not explicitly attributed positions in space. Nevertheless, the one-dimensional ordered sequence of numbers is associated with a one-dimensional imaginary physical line, according to the characteristics of the experiment.

The original interpretation of the SNARC states that small numbers are represented to the left and large numbers to the right (see Section 3.4.1). In other words, lower magnitudes are anchored to the left hemispace and higher magnitudes to the right hemispace, even though there is no explicit SNL. This interpretation can be reformulated by stating the existence of a horizontal as-if SNL with lower magnitudes to the left and higher magnitudes to the right.

The origins of the SNARC are still debated. Although the original interpretation has been proven to be too naive (see Section 3.4.4), it serves the purpose of leading the review of associations between numbers and space without SNLs. The question which I seek to answer is whether the MNL is permanently anchored to an as-if SNL, either by innate neural connections or by long-term associative learning, or if short-term associations between the MNL and different as-if SNLs are possible. To do so, I review number-based decisions where spatial attributes are task-irrelevant.

3.4.1. The SNARC effect

Dehaene et al. (1993) design a simple psychological experiment intended to verify the triple-code model of numerical processing. Subjects perform parity judgments on single- or double-digit numbers appearing on a screen, e.g. pressing a left button for even numbers and a right button for odd numbers. According to the triple-code model, parity and magnitude are processed independently, so that response times should be similar for all numbers; this is precisely the result they get.

However, because they have to counterbalance the responses, they assign the buttons randomly on each experimental block. Much to their surprise, they find an interaction between side of response and number magnitude. Smaller numbers have faster responses with the left button and larger numbers have faster responses with the right button, thus producing a characteristic negative slope (Fig. 3.10). They interpret this result as indicating that numerical magnitude is automatically processed even if task-irrelevant and somehow interacts with the spatial codes of the responses. They call this phenomenon the SNARC effect.

Dehaene et al. (1993) explore different aspects of the SNARC effect in a series of modified versions of the parity judgment task. First, they show that the associations of
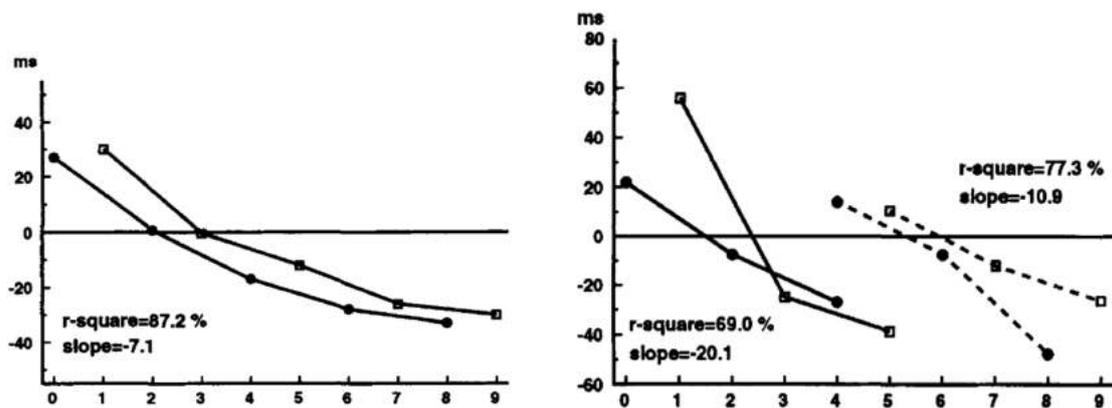
Figure 3.10: The original SNARC effect (Dehaene et al., 1993). Left: original SNARC, obtained as the difference in response times between the left and right buttons. The negative slope indicates a left-to-right mapping. Right: relative SNARC, with two different intervals used.

numbers and space are relative to the numerical interval tested: the numbers 4 and 5 are large in the interval 0-5, and therefore facilitate right responses, but they are small in the interval 4-9, and therefore facilitate left responses (Fig. 3.10). Analogously to the results of Wallace (1971) with the Simon effect, the SNARC effect persists when subjects cross their hands. Hence, the MNL is mapped to an as-if SNL that spans the space between the responses, regardless of the body part used to execute the actions.

The influence of reading direction

Importantly, these initial experiments exclusively involve French participants. Dehaene et al. (1993) hypothesize that the direction of reading and writing could be responsible for the direction of the SNARC effect, since the French language involves reading and writing from left to right. They test a group of Iranian participants living in France who read and write from right to left. They find no SNARC effect for the entire group; however they find a positive slope, i.e. a reversal of the SNARC, for participants with less than two years of living in France, therefore less influenced by the Western writing system.

The cultural origin of the SNARC effect is further studied by Shaki et al. (2009). They are interested in distinguishing whether the number-space mapping matches the direction of reading words or numbers. Shaki et al. recruit three different types of subjects: Canadians, who read both words and numbers from left to right; Palestinians, who read both words and numbers from right to left; and Israelis, who read words from right to left but numbers from left to right. They find a typical negative slope for Canadian participants, i.e. a conventional SNARC effect, and a positive slope for Palestinian participants, i.e. a reversed SNARC effect, but no SNARC effect for Israeli participants. These results show that both the reading of numbers and the reading of text influence the mapping of numbers to space.

However, a vertical SNARC effect has also been reported, with smaller numbers facilitating the bottom response and larger numbers facilitating the top response (Winter et al., 2015). Reading and writing cannot be at the origin of the vertical SNARC, although the experience of stacked objects may be, i.e. the association of *more* with *higher*.

SNARC in infants and animals

Following the results presented in Sections 3.2.3 and 3.3.2, I review the evidence for an as-if SNL in infants and non-human animals. If the associations between numbers and space are acquired from experience, then newborn humans or animals may not show any spatial biases in number-based tasks.

This is precisely what De Hevia et al. (2017) test with 0- to 3-day-old neonates. First, infants are habituated to a repeated sequence of sounds while one line appears on a screen. During the test trials, the number of sounds of the sequence changes, and two different lines appear on the screen, one to the left and one to the right. For example, some infants are habituated with 18 sounds and a long line, and tested with 6 sounds and short lines; others are habituated with 6 sounds and a short line, and tested with 18 sounds and long lines. De Hevia et al. (2017) find longer looking times towards the line on the left when the number of sounds decreases and towards the line on the right when the number of sounds increases. These results indicate a horizontal as-if SNL going from left to right.

Animals have an advanced sense of space (see Section 2.1) as well as basic numerical abilities. To test whether the two are related, Rugani et al. (2015) perform a simple task with newborn chicks. The animals are trained to find food behind a panel displaying an array of dots of a specific target numerosity. They are then tested with two identical panels side-by-side with the same numerosity; on half of the tests the numerosity is smaller than the target and on the other half it is larger. First, they train chicks with the numerosity 5 and test them with 2 vs. 2 and 8 vs. 8. They find significant preference for the left panel with the numerosity 2 and for the right panel with the numerosity 8, indicating an association of small numbers with the left space and of large numbers with the right space. To find out whether the effect is still present for larger numerosities, they repeat the experiment with the target 20 and the tests 8 vs. 8 and 32 vs. 32; they find the same results.

Even though the numerical abilities of chicks are undisputed, the reason for the asymmetric choices in the experiments of Rugani et al. (2015) are debated. The authors claim that their results reveal an evolutionary predisposition to map numbers onto space embodied in neural systems, probably through brain asymmetry. However, these results were not replicated in a recent study with rhesus monkeys (Beran et al., 2019).

A connectionist model of the SNARC effect

Gevers et al. (2006) modify a previous connectionist model of number recognition (Verguts et al., 2005) to reproduce behavioral data from the SNARC effect. They do so by including an intermediate layer with three attributes: magnitude, either small or large, parity, either odd or even, and an extra field for arbitrary number-response mappings. The output of the model consists of two nodes for the left and right responses. The model of Gevers et al. (2006) reproduces the SNARC effect for both parity and magnitude judgments.

Moreover, they also conduct a human SNARC experiment to study the dependence of the effect with response time. They find an increasing SNARC effect as responses take longer, which is equivalent to a positive slope in a delta plot (see Section 2.3.1). The computational model of Gevers et al. (2006) is able to capture this behavior.

3.4.2. Other spatial-numerical associations

In terms of the number lines defined previously, the SNARC effect can be interpreted as a difference in response times because the MNL is mapped onto an as-if SNL. The task-irrelevant activation of the numerical magnitude on the MNL leads to an isomorphic activation on an as-if SNL which interferes with the spatial representations of the responses. However, the SNARC effect is not the only effect that reveals

spatial-numerical associations. In this Section, I look into three other phenomena which show how the MNL is mapped onto an as-if SNL.

Attentional SNARC

A closely related phenomenon to the SNARC effect is reported by Fischer et al. (2003). They perform a Posner cueing task, where participants have to press a button on the same side of a flash, following a task-irrelevant number cue. Their results indicate that smaller numbers lead to faster responses to the left, while larger numbers lead to faster responses to the right. In line with previous results from the Posner cueing task (Posner, 1980), they conclude that seeing a number leads to an attentional shift towards the hemispace corresponding to the number's position in the MNL. This effect receives the name of attentional SNARC.

Operational momentum

One particularly interesting numerical effect is the so-called operational momentum. McCrink et al. (2007) design an experiment where participants watch several clips of numerosity additions and subtractions, and have to quickly estimate the result of the arithmetical operation. For additions, participants show a tendency to overestimate the results; for subtractions, they show a tendency to underestimate the results. McCrink et al. (2007) identify the difference between the two cases as a being similar to a physical momentum: on the MNL, adding one numerosity to another implies a movement from the first one towards larger magnitudes, and the over-estimation is attributed to a tendency to continue said movement, i.e. a momentum. Likewise, subtracting two numerosities implies a movement towards smaller magnitudes and under-estimating the result. The term operational momentum refers to the momentum resulting from performing an approximate arithmetic operation on the MNL. Pinhas and Fischer (2008) repeat the experiment with symbolic operations, but instruct subjects to point to their responses on a number line estimation task. They find that subtractions lead to left biases and additions to right biases. Using finger-tracking, Pinheiro-Chagas et al. (2017) show that the processing of the two operands is serial: the finger initially moves towards the first operand, and the trajectory is then modified according to the sign of the operation and the magnitude of the second operand. They find that the operational momentum originates when incorporating the sign of the operation.

Patients with hemispatial neglect

The line bisection task briefly mentioned in Section 3.3.2 has been a particularly useful tool to test patients with hemispatial neglect. This is an acquired neuropsychological disorder, typically due to lesions to the right cerebral hemisphere which affects the conscious perception of the left hemispace. Bisiach and Luzzatti (1978) report two patients who, when asked to describe a well-known scenario like the Piazza del Duomo in Milan, omit left-sided details from different perspectives. Patients with hemispatial neglect are unable to reproduce complete drawings, for example ignoring the left side of a clock face. The line bisection task is used to measure the degree of the lesion in these patients: rather than bisecting the line in the middle point, they shift it to the right as if ignoring the left half of the line.

Zorzi et al. (2002) implement a task inspired by the line bisection which they call the number bisection task. A patient is given two numbers which form a numerical interval and asked to indicate quickly and without calculations which number lies exactly halfway. For example, for the interval 3-7, the correct response is 5. Quite astonishingly, Zorzi et al. (2002) find that the responses of patients with hemispatial neglect are consistently deviated
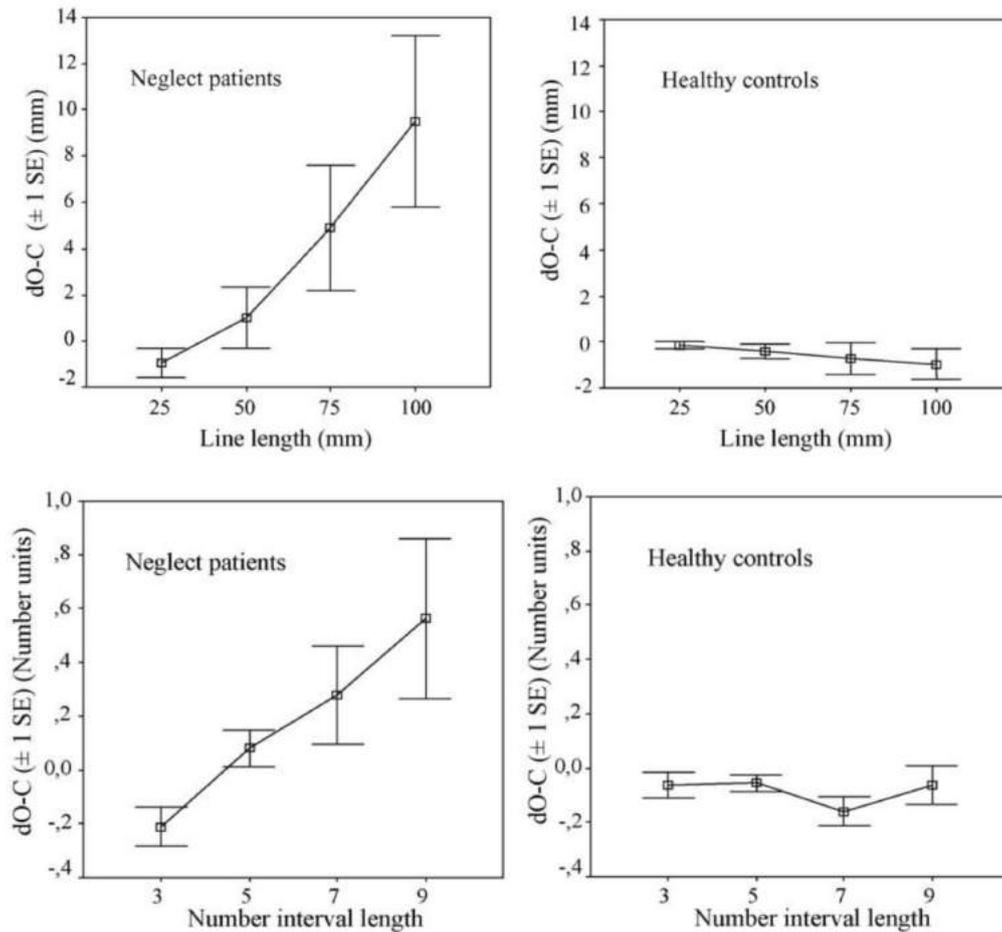
Figure 3.11: Results from patients with hemispatial neglect (Zorzi et al., 2006). Top: line bisection task, bottom: number interval bisection task. In both cases the deviation from the middle increases with the interval length for patients with hemispatial neglect but not for controls.

towards the larger numbers, as if they ignored the lower half of the number interval. In the example, a patient with neglect would say that the number exactly in the middle of 3 and 7 is 6.

The similarity between the patterns of line bisections and number bisections from patients with hemispatial neglect is remarkable, with biases towards the right and towards larger numbers which increase with line length and with numerical interval. This is not observed with letters or months (Zorzi et al., 2006).

Because the deviations are relative to the numerical interval tested, these results are interpreted as indicating that patients with hemispatial neglect are unable to directly access the lower end of the MNL, in much the same way that they are unable to access the representation of their left hemispace (Priftis et al., 2006; Umiltà et al., 2009).

3.4.3. Overlapping neural codes in parietal cortex

So far, I have discussed number-space associations in behavioral experiments with humans and non-human animals, as well as patients with lesions in the parietal cortex. I presented the idea of an as-if SNL, i.e. a spatial mapping of the MNL as if numbers were displayed on a SNL. But is there any neural evidence to back such associations between mental representations? The list of brain regions involved in both number and space
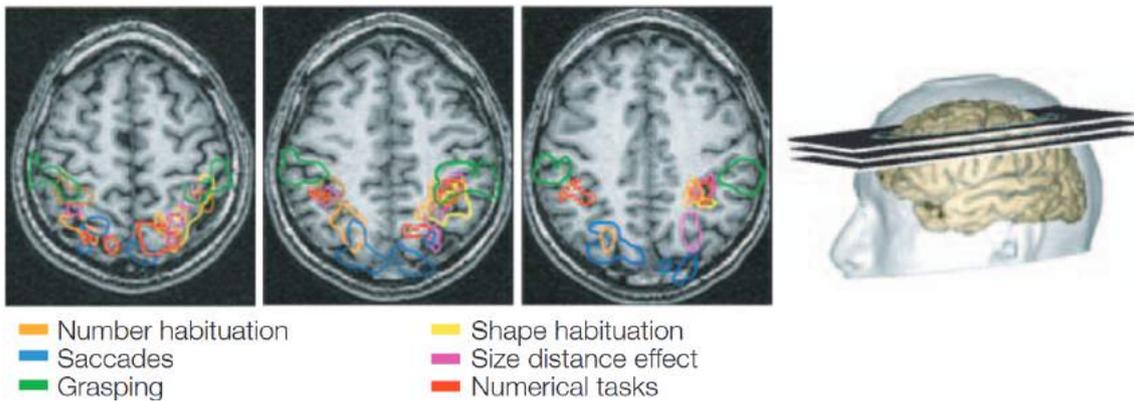
Figure 3.12: Location of spatial and numerical processing in the human brain (Hubbard et al., 2005). Overlaps can be identified between number-related and space-related tasks, in a region coinciding with the IPS.

representations (Sections 3.2.4 and 2.1) include common areas. In particular, the IPS in the PPC emerges as a relevant hub where spatial, numerical, and attentional codes are integrated.

Hubbard et al. (2005) discuss the roles that the different areas within the IPS may play in number-space interactions. From an extensive review of human neuroimaging and animal electrophysiology studies, they identify the monkey area VIP, and its putative human homologue, as being involved in spatial-numerical mappings. Both human and animal studies reveal that numerical quantities are coded in the fundus of the IPS, a region identified in monkeys as roughly coinciding with area VIP. VIP is also responsible for coding multisensory spatial information about objects in close proximity. The multiple representations of space in and around the IPS appear to be closely related to number-space associations. They suggest that number representations may have emerged from the neuronal recycling of areas responsible for spatial representations. For example, regarding the similarities between approximate arithmetic operations and the transformation from eye- to head-centered coordinates, Hubbard et al. state that "parietal mechanisms that are thought to support spatial transformation might be ideally suited to support arithmetic transformations as well" (Hubbard et al., 2009, p. 238).

However, recent findings challenge the view of overlapping neural codes for number and space (Kanayet et al., 2018). Subjects see number digits and words and also perform a number line estimation task, all inside a high-resolution fMRI scanner. The researchers analyze the differences between conditions and find a functional dissociation between two areas of the IPS. The anterior IPS is activated for number processing and recognition, i.e. the neural correlate for the MNL, whereas the posterior IPS is activated when participants prepare the positioning of the number on the spatial line, i.e. the neural correlate for the SNL.

Two alternative proposals emerge from the analysis of neuroimaging and electrophysiological data. In the words of Kanayet et al.: "one interpretation (which we will call the positional view) states that mental representations of numbers are intrinsically organized by position, potentially as a result of the brain recycling the neural architecture used to represent space [...]. An alternative interpretation, which we will call the magnitude view, holds that numbers are represented as an approximate magnitude [...], which is not intrinsically positional, but which can be mapped onto position in an experience- and context-dependent fashion by providing a position reference." (Kanayet et al., 2018, p. 200).

I believe the latter, i.e. the magnitude view, to be a more parsimonious interpretation.

The magnitude representation of numbers in the anterior IPS and the spatial representation of numbers in the posterior IPS are very strongly related, with monosynaptic connections between different regions of the IPS. Even though there appears to be no direct overlap of the neural codes, these strong connections between the neural regions responsible for the MNL and the SNL could lead to automatic activations of spatial codes in tasks not involving spatial mappings of numbers, e.g. the SNARC effect. This is what I have defined as an as-if SNL.

'

3.4.4.   The role of short-term associations

One of the original experiments from Dehaene et al. (1993) reveals that the mapping of numbers onto space responsible for the SNARC effect is relative: numbers are small or large with respect to the numerical interval of the experiment. Therefore, it appears unreasonable to assume associations between numbers and space are fixed and exclusively of long-term nature. Here, I list a few experiments with short-term reversals of the SNARC to highlight the role of working memory in number-space mappings.

Bächtold et al. (1998) design a SNARC task in which subjects are asked to perform number comparisons in two different spatial frames. First, subjects have to indicate whether a number is "shorter or longer than 6cm", i.e. a typical SNL; they find a conventional SNARC effect, with faster responses to the left for smaller numbers and to the right for larger numbers. Next, subjects have to indicate whether a number is "earlier or later than 6 o'clock", i.e. with respect to a clock face, an atypical SNL in which smaller numbers are represented to the right and larger numbers to the left; the results indicate a reversal of the SNARC effect, in line with a right-to-left as-if SNL. Rather than considering the possibilities of multiple permanently-mapped MNLs, each with a specific spatial orientation, the results of Bächtold et al. (1998) support the idea of magnitude representations in a single MNL that are implicitly mapped onto space according to the instructions of the task.

Similar short-term reversals of the SNARC effect can be obtained from directional ocular movements. Ranzini et al. (2016) report a relation between spatial-numerical associations and attention by having subjects perform a parity judgment while tracking a moving dot with their gaze. They find that large numbers are processed slower during leftwards ocular motions than during rightwards ones. Shaki and Fischer (2008) test bilingual Russian-Hebrew subjects. After reading Cyrillic texts, from left to right, a conventional SNARC is found; however, when subjects read Hebrew texts, from right to left, the SNARC effect is reduced. The researchers then repeat the procedure with participants listening to texts instead of reading, and they find no differences. They conclude that the spatial direction of scanning visual materials contributes to the short-term spatial mapping of numbers. The opposite has also been observed: Loetscher et al. (2010) ask participants to say random numbers while they track their eye movements, and find a strong correlation between the difference between subsequent numbers and the change in eye position.

Interestingly, the SNARC effect can reveal the origin of the spatial coding of the as-if SNL. Gevers et al. (2010) design a series of experiments with different types of spatial codes that reveal different number-space associations. In one in particular, they add the written tags "LEFT" and "RIGHT" to the two response buttons, on half of the trials congruently and on the other incongruently. They find a stronger interaction between numbers and tags than between numbers and position, i.e. smaller numbers mapped to the LEFT tag and larger numbers mapped to the RIGHT tag, not to the left and right buttons. In their own words, the verbal-spatial coding scheme is dominant over the visuospatial coding. According to Gevers et al. (2010), these results reveal that number representations are not
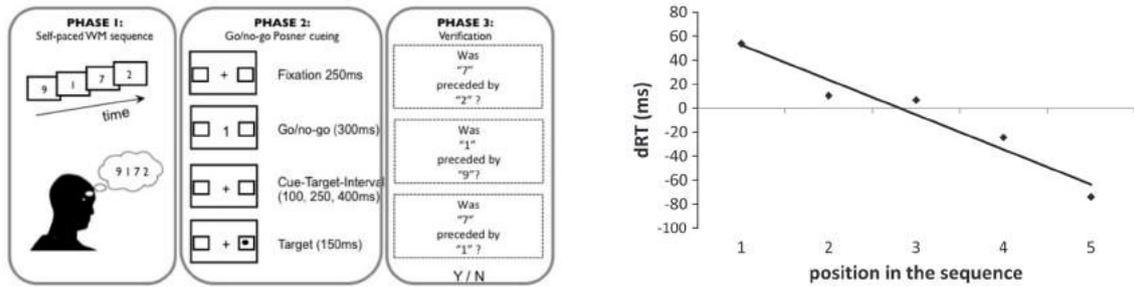
Figure 3.13: The role of working memory in the SNARC effect. Left: attentional SNARC task with working memory sequence (van Dijck et al., 2013). Participants are instructed to maintain a sequence of numbers in working memory and then perform a Posner cueing task with those same numbers and others. They are later tested to verify that they remember the original sequence of numbers. Right: response time differences as function of the position in the working memory sequence (van Dijck and Fias, 2011) which indicates that the SNARC effect depends on serial order coding in working memory.

merely activating visuospatial representations. In accordance with this conclusion, Mapelli et al. (2003) perform a SNARC and Simon task simultaneously, by presenting numbers for a parity judgment to either side of the screen, and find non-interactive additive effects.

One line of work that has proven particularly resourceful is the study of the role of working memory in spatial mappings. In an experiment by van Dijck and Fias (2011), subjects have to memorize a sequence of 5 numbers, and then perform a SNARC task. Rather surprisingly, they find that, regardless of their numerical magnitude, it is their position in the sequence that determines the hemispatial facilitation, i.e. the first numbers in the sequence have faster responses with the left button and the later numbers have faster responses with the right button. This and other similar experiments (e.g. Fias et al., 2011; van Dijck et al., 2013) lead the authors to suggest a theoretical framework where serial order in working memory is responsible for number-space associations, thus rejecting the ideas of long-term and even innate mappings (Fias and van Dijck, 2016).

3.5. A proposal for number-space mappings

Having reviewed some of the main experimental findings and existing accounts regarding the spatial representation of numbers, I can now present my own theoretical view, as well as computational models to support it. I begin this Section with a simple model of number comparisons. Although this does not include any spatial mappings, reproducing the numerical distance effect is an important first step for any model of numerical cognition. The same model is then adapted to include the spatial component of two lateralized buttons, which gives rise to the SNARC effect.

3.5.1. An LCA approach to the numerical distance effect

The numerical distance effect refers to a decrease in response time when comparing two numbers as a function of their numerical difference. This phenomenon emerges in multiple different tasks, but here I focus on a number comparison to a target (Hinrichs et al., 1981; Dehaene et al., 1990). Subjects are given one target per experimental block, e.g. the number 55, and have to press one button if a two-digit number shown on a screen is lower and another button if it is higher. The results from this task are in line with the results from other number comparison tasks (Moyer and Landauer, 1967; Sekuler and Mierkiewicz, 1977): higher response times as numbers get closer to the target (Fig. 3.14).

The numerical distance effect depends exclusively on a single attribute: the magnitude of the number displayed on the screen. Therefore, this effect can be modeled with a conventional single-attribute LCA model (Usher and McClelland, 2001) with two accumulators, *lower* and *higher*. The dynamics of such a system are described in Section 1.2.1.

First, let $\{s_n\}$ be a set of numerical magnitude vectors, where n takes values between 11 and 99. These are normalized, two-dimensional vectors that describe an arc between the two extremes of the numerical interval. The angle of each vector can be sampled from a Gaussian distribution to account for the overlapping representations of the MNL.
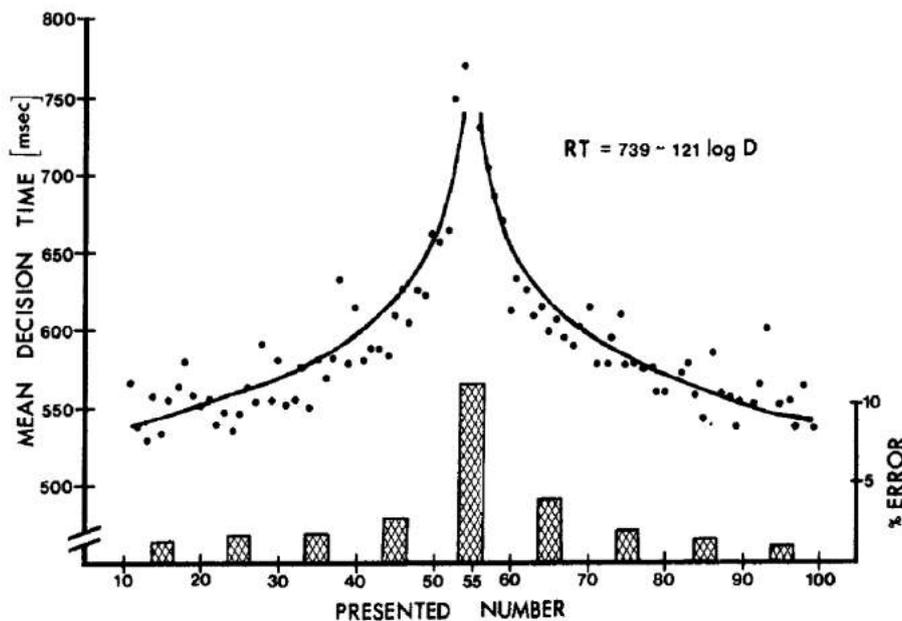
Figure 3.14: Response times during a number comparison task, where numbers are compared to 55 (Hinrichs et al., 1981). Response times are fit with a logarithmic dependence of numerical distance.

Numbers lower than 55 must be associated with the *lower* accumulator and numbers higher than 55 must be associated with the *higher* accumulator. This can be done by means of a linear associative memory given by

$$\boldsymbol{M} = \eta_{lower} \ \boldsymbol{r}_{lower} \ \boldsymbol{s}_{lower}^T + \eta_{higher} \ \boldsymbol{r}_{higher} \ \boldsymbol{s}_{higher}^T \tag{3.1}$$

where η_{lower} and η_{higher} are the scaling rates, $\boldsymbol{s}_{lower}$ and $\boldsymbol{s}_{higher}$ are the stored number vectors and $\boldsymbol{r}_{lower}$ and $\boldsymbol{r}_{higher}$ are the response vectors corresponding to each accumulator. The stored number vectors are abstract representations inside or outside of the numerical interval that act as anchors; it is assumed that this memory is developed during an initial training phase in which these abstractions are generated. The integration process follows (1.8), that is

$$d\boldsymbol{x} = [\boldsymbol{M}\boldsymbol{s}_n - \boldsymbol{D}\boldsymbol{x}] \frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{3.2}$$

and finishes when one of the accumulators reaches a threshold value. The corresponding response is selected. When a number n is presented on the screen, the magnitude representation $\boldsymbol{s}_n$ is automatically generated (see Appendix A) and the output of the matrix serves as input into the accumulators:

$$\boldsymbol{M}\boldsymbol{s}_n = \eta_{lower} \ \boldsymbol{r}_{lower}\langle \boldsymbol{s}_{lower}, \boldsymbol{s}_n \rangle + \eta_{higher} \ \boldsymbol{r}_{higher}\langle \boldsymbol{s}_{higher}, \boldsymbol{s}_n \rangle \tag{3.3}$$

In order for this model to correctly reproduce the results from a number comparison task, numbers lower than the target must activate the lower response, and numbers higher must activate the higher response. In particular, the parameters of the model must satisfy the condition that, if the number presented is the target itself, then activation for both responses must be equal, hence

$$\frac{\eta_{higher}}{\eta_{lower}} = \frac{\langle \boldsymbol{s}_{lower}, \boldsymbol{s}_{target} \rangle}{\langle \boldsymbol{s}_{higher}, \boldsymbol{s}_{target} \rangle} \tag{3.4}$$

This is readily satisfied by making the scaling rates equal ($\eta = \eta_{lower} = \eta_{higher}$) and the stored number vectors $\boldsymbol{s}_{lower}$ and $\boldsymbol{s}_{higher}$ symmetric with respect to the target. However, the symmetry must be in the metric of the corresponding number representation. Symbolic tasks, such as the experiment performed by Hinrichs et al. (1981), can be modeled with a linear scale. Alternatively, non-symbolic tasks are better captured with logarithmic compression (see Section 3.2.2).

In either case, the number line is represented along an arc between the extremes of the numerical interval, where the angles between vectors are proportional to their numerical difference. One can arbitrarily assign one of the stored vectors to the extreme of the interval that is furthest from the target number, in the metric of the number line. The other stored vector is assigned to the number that lies symmetrically to the first one with respect to the target.

First, we can look at a model of the experiment by Hinrichs et al. (1981). As mentioned before, this is a symbolic task and so I use a linear number line representation with target 55 and extremes 11 and 99 (see Appendx B for details on the numerical simulations). Fig. 3.15 shows the response times as a function of the number presented on the screen. As numbers get closer to the target, response times increase. The symmetry from these numerical simulations stems from the symbolic nature of the task. Unlike some previous works (e.g. Sekuler and Mierkiewicz, 1977), Hinrichs et al. (1981) suggest that a logarithmic dependence of response times with respect to numerical distance is a better fit. Here, it is important to remember that the experiment of Henrics et al. involves two-digit numbers while Sekuler and Mierkiewicz (1977) use single digits. Our results support the logarithmic
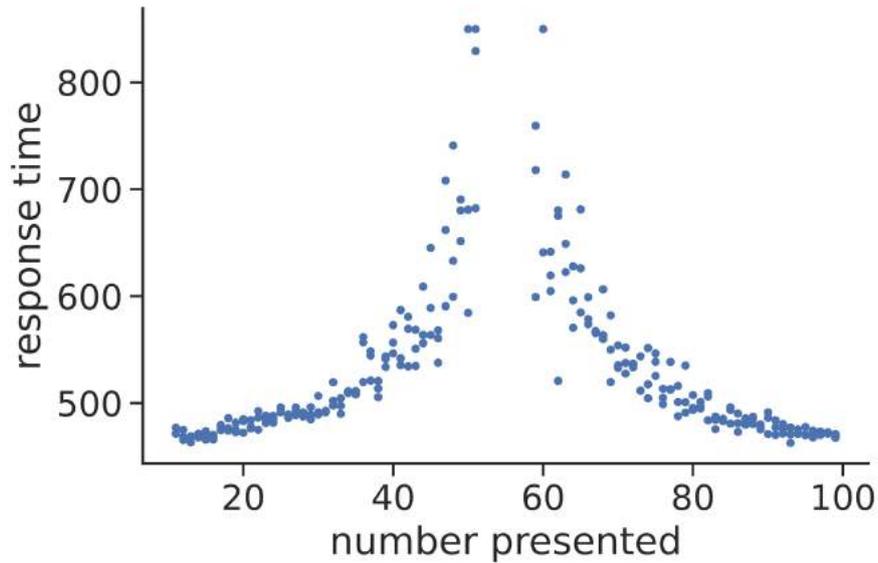
Figure 3.15: Simulation of a number comparison task, with a linear mapping of numbers. Response times are higher for numbers closer to the target number.

dependence of the numerical distance effect, at least in a segment of the numerical interval (Fig. 3.16).

Next, we can see the difference of using a compressed scale. I repeated the previous simulations with a logarithmic representation of numbers, akin to the MNL. Fig. 3.17 shows the response times of the same simulation when plotted on a linear scale and on a logarithmic scale. One can quickly identify that the former is skewed, while the latter is symmetric. The same results are obtained in behavioral experiments and physiological
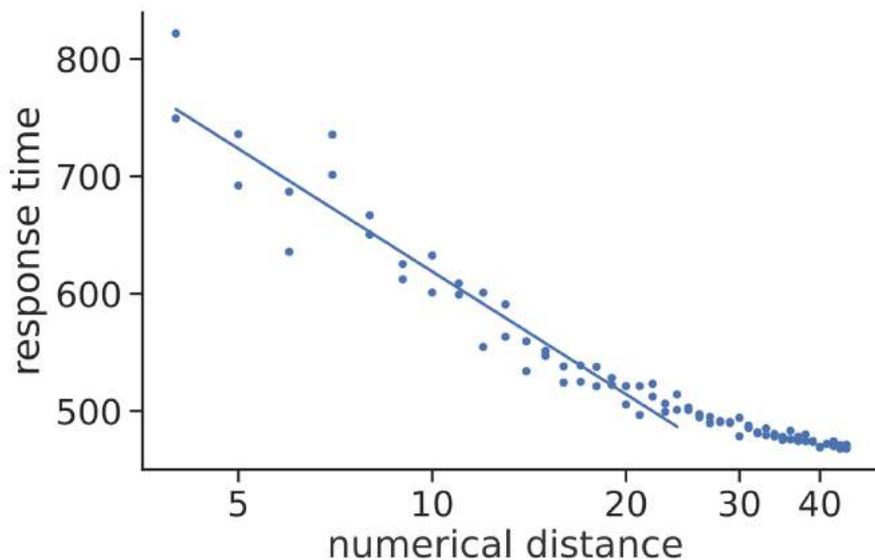
Figure 3.16: Regression of numerical distance effect. Response times are plotted as a function of the numerical distance on a logarithmic scale. A linear relation is observed for numerical distances lower than 25, whereas response times show little change for larger numerical distances.
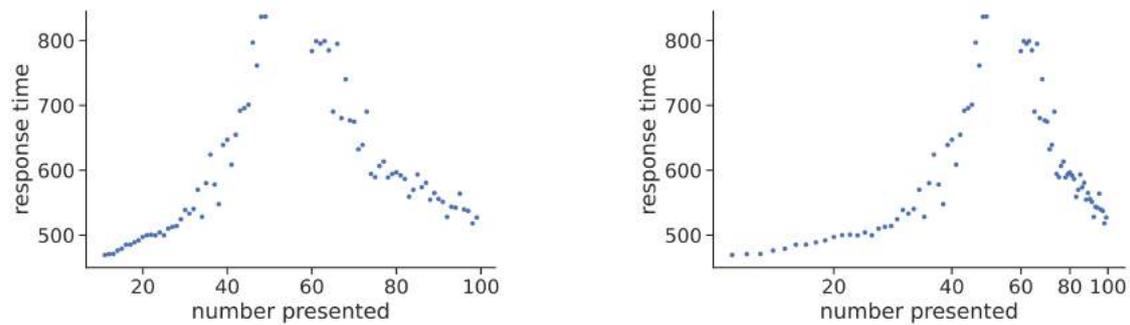
Figure 3.17: Simulation of a numerosity comparison task, obtained with a logarithmic mapping of numbers. An asymmetry is observed when plotting on a linear scale (left), which disappears when plotting on a logarithmic scale (right). Similar results are obtained at the neural and behavioral level (Dehaene, 2003).

recordings (Dehaene, 2003; Nieder et al., 2002). These LCA simulations support the argument that the MNL has logarithmic compression.

In conclusion, this LCA approach to number comparisons is capable of reproducing two well-known results from the field: the numerical distance effect for symbolic tasks and the logarithmic compression of non-symbolic representations. The model can therefore be used as a foundation to capture the SNARC effect. Before, I must state my theoretical views on the spatial representation of numbers.

3.5.2. Three levels of number-space associations

Given the quantity and diversity of experiments intended to constrain number-space mappings to a single cause, it is apparent that there are different levels of associations which play different roles with different degrees of influence. These span from neural connections between numerical and spatial representations to short-term associations from serial order working memory.

In a recent review on the relation between mathematical and spatial skills, Hawes and Ansari (2020) identified four non-mutually exclusive accounts: a spatial representation of numbers account, a shared neuronal processing account, a spatial modeling account, and a working memory account. The spatial modeling account refers to how mathematical operations can be visualized in a "mental whiteboard"; it is not particularly relevant to the SNARC effect and other related phenomena.

The remaining three accounts identified by Hawes and Ansari (2020) broadly coincide with what I refer to as the three levels of number-space associations. These are non-mutually exclusive potential origins for the spatial representation of numbers, likely overlaid on top of each other through ontogenic development.

Shared neural processing

An increasing corpus of behavioral studies in infants (De Hevia et al., 2014; De Hevia et al., 2017), animals (Rugani et al., 2007; Rugani et al., 2015; Drucker and Brannon, 2014), and patients with hemispatial neglect (Zorzi et al., 2002; Priftis et al., 2006; Umiltà et al., 2009), reveals that spatial-numerical associations may have an origin in shared neural processing.

Both numbers and space are known to be represented in the IPS (Hubbard et al., 2009), likely due to the neuronal recycling of brain areas originally evolved for interactions with the physical world that can also support basic mathematical abilities (Dehaene and Cohen, 2007). The initial suggestion of overlapping neural codes in humans (Hubbard et al., 2005)

is challenged by the use of more fine-grained neuroimaging which reveals two separate codings, with number magnitudes represented in the anterior IPS and spatial mappings of numbers in the posterior IPS (Kanayet et al., 2018).

My view is that the shared neural processing of magnitudes and space is the first level for number-space associations. This is an innate mapping, available from birth, due to the strong connections between different areas of the IPS. The MNL likely has a neural substrate in the human anterior IPS and monkey area VIP, while the SNL may be represented in the human posterior IPS and monkey area LIP. The monosynaptic connections between these regions can explain why an as-if SNL is deployed in non-spatial tasks such as the SNARC effect. This hypothesis will require more advanced neuroimaging techniques in humans, since electrophysiological studies in animals are restricted by the limited evidence of number-space mappings (Beran et al., 2019).

Long-term associations

A second origin for number-space mappings is related to the long-term associations from experience, mainly due to formal education. From the moment they are born, children interact with the world around them and learn from these interactions. Their elementary innate numerical abilities (Izard et al., 2009) improve rapidly, both with (Halberda et al., 2008) and without formal education (Pica et al., 2004). Illiterate indigenous populations are able to develop number-space mappings after training (Dehaene et al., 2008), while the use of number lines enhances the number sense and mathematical abilities of children (Newcombe et al., 2018; Lourenco et al., 2018). The direction of reading and writing determines the direction of the SNARC: Westerners, who read from left to right, display a negative SNARC slope (Dehaene et al., 1993; Shaki et al., 2009), while Palestinians, who read from right to left, display a positive SNARC slope (Shaki et al., 2009). According to these results, the as-if SNL for a SNARC task is mapped along the direction of reading. The visual experience of correlations of quantity and height also accounts for the vertical SNARC (Winter et al., 2015).

I propose long-term associations act as a second level of number-space mappings, built on top of the basal neural ones. These associations stem from interactions with the physical world and with others. The repeated correlations of *more objects* with *up*, as well as *more text* with *right* or *left* depending on the culture, shape these associations. Symbolic mathematics also play an important role. Starting with the rudimentary SNL of Wallis (1685), the explicit mapping of ordered numbers on lines, which are used to learn about ordinality, cardinality, and arithmetics, further sediment these culture-dependent associations, while simultaneously reshaping number line estimations from compressed to linear ones (Siegler and Opfer, 2003; Sella et al., 2020). Long-term mappings of numbers onto space are revealed in SNARC, attentional SNARC and operational momentum tasks. Overtly or not, participants in these experiments draw upon the relations stored in memory. These associations are represented as a permanent mapping of the MNL to an as-if SNL which coincides with the SNLs most frequent in each individual's experience.

Working memory

Lastly, a third level of number-space mappings is possible with short-term associations in working memory. These override the basal and long-term associations by means of specific task-related instructions, such as modifications to the spatial layout of numbers (Bachtold et al., 1998) and of responses (Gevers et al., 2010), reversals of reading directions (Shaki and Fischer, 2008), and serial order coding in working memory (van Dijck and Fias, 2011; Fias et al., 2011). The working memory account is further supported by the need for participants to perform several practice trials to fully understand and assimilate the

instructions of any experiment.

Some authors have suggested that working memory alone is responsible for spatial-numerical associations (Fias and van Dijck, 2016), but I believe this to be excessive. Instead, I suggest number-space associations are ruled by working memory whenever a task implements a mapping which includes mismatches with respect to the previously stored one. That is, the as-if SNL which results from neural and cultural mappings can be momentarily altered if it does not correspond with the task at hand or if another as-if SNL can improve performance.

Multi-level attentional interactions

As a final note, it is relevant to highlight that the three levels of spatial-numerical associations in this theoretical model are overlaid such that one can access a lower level with certain experimental conditions. Here, attention is of key importance. Participants using working memory to solve a task also rely on long-term memory (Loetscher et al., 2010; Guida et al., 2018); long-term associations are also disrupted when attentional resources are limited (Anobile et al., 2012). Attention and serial order working memory may have neural correlates in the IPS (Majerus et al., 2006; Attout et al., 2014).

I suggest that some form of Hebbian learning might be involved in the spatial representation of numbers. In a single experiment, attention and working memory mediate the associations between magnitude codes in the anterior IPS and of the spatial codes in the posterior IPS. The repeated experience of attending to certain numbers at certain positions leads to the repeated simultaneous firing of the same populations of neurons in those two areas in the IPS, thus leading to the establishment of a memory. Given the strong synaptic connectivity in the IPS, a numerical representation could activate an associated spatial representation, even if the latter is task-irrelevant.

3.5.3. An application of MLCA to the SNARC effect

Regardless of whether the number-space mapping uncovered by the SNARC arises from neural, cultural or working memory associations, it is of interest to build a computational model that reproduces the behavioral data. To do so, I first identify that the SNARC effect involves two different attributes of a single stimulus, i.e. the number appearing on a screen: parity, which is task-relevant, and magnitude, which is task-irrelevant. In terms of SRC, the SNARC has a dimensional overlap between the task-irrelevant stimulus attribute and the responses, and it is therefore a type 3 ensemble (Kornblum et al., 1990; see Table 2.1). The SNARC effect is similar to the Simon effect in that incongruencies originate from the task-irrelevant attribute, but it is not *per se* a Simon effect (Mapelli et al., 2003).

Since the SNARC effect involves two attributes, I use the MLCA model developed in Section 1.3 with two accumulators. In addition, and unlike the model of the Simon effect, this one can have a single route of accumulation, because the delta plots have positive slopes (Gevers et al., 2006; see Section 2.3.1).

I begin by defining the two sets of attributes which form a stimulus. For numerical magnitude, I use the same representations from Section 3.5.1 to reproduce the numerical distance effect. Each number n is assigned a normalized vector s_n which lies on an arc between the two extremes of the numerical interval of the task at hand. Here, the interest is to capture the behavioral data from the original SNARC effect discovery (Dehaene et al., 1993) with a numerical interval $[1, 8]$, with anchors s_{low} and s_{high} which can be set to coincide with the extremes of the numerical interval.

On the other hand, parity is an abstract concept which emerges from the visual arabic number system of the triple-code model (Dehaene, 1992). Since parity is a purely symbolic

attribute shown to be independent of magnitude (Dehaene et al., 1993), I can assign two orthonormal vectors to the two possible parities: $\boldsymbol{p}_{even}$ and $\boldsymbol{p}_{odd}$.

Next, one can build the context-dependent associative memory $\boldsymbol{M}$ that produces the readouts of the attribute representations. Here, there is no interest in the training phase of an experiment. Hence, I can assume that the subject fully understands the instructions of the task and will therefore only store associations from each parity to the correct response. That is, if the instruction is to press the left button upon seeing an even number, the memory only stores the association of even numbers to the left accumulator. Such a model represents decisions based on accuracy rather than on speed. However, the same reasoning does not apply to the associations of magnitudes and responses. Even after training, that is, after a working memory mapping of the MNL to a task-specific as-if SNL, previous number-to-space mappings may remain active. It is therefore best to consider two sets of associations with different parameters, similarly to what was done to model the automatic process of the Simon effect in (2.8).

The dynamics follow (1.14), that is

$$\boldsymbol{dx} = [\boldsymbol{M}\,(\boldsymbol{s}_n \otimes \boldsymbol{p}) - \boldsymbol{D}\boldsymbol{x}]\,\frac{dt}{\tau} + \boldsymbol{\xi}\sqrt{\frac{dt}{\tau}} \tag{3.5}$$

where $\boldsymbol{M}$ is the context-dependent associative memory for the corresponding condition and $\boldsymbol{s}_n$ and $\boldsymbol{p}$ are the vectors that represent the number shown on the screen. Given the instruction for what button to press for each parity, such as *even-left* and *odd-right*, the context-dependent associative memory $\boldsymbol{M}_{EL,OR}$ is given by

$$\begin{aligned}
\boldsymbol{M}_{EL,OR} = \;& \eta\,\boldsymbol{r}_{left}\,({\boldsymbol{s}_{low}}^T \otimes {\boldsymbol{p}_{even}}^T) + \\
& \eta'\,\boldsymbol{r}_{right}\,({\boldsymbol{s}_{low}}^T \otimes {\boldsymbol{p}_{odd}}^T) + \\
& \eta'\,\boldsymbol{r}_{left}\,({\boldsymbol{s}_{high}}^T \otimes {\boldsymbol{p}_{even}}^T) + \\
& \eta\,\boldsymbol{r}_{right}\,({\boldsymbol{s}_{high}}^T \otimes {\boldsymbol{p}_{odd}}^T)
\end{aligned} \tag{3.6}$$

where $\eta > \eta'$ are the scaling rates and $\boldsymbol{r}_{left}$ and $\boldsymbol{r}_{right}$ are the response vectors corresponding to each accumulator. The alternative instruction, *odd-left* and *even-right*, has a context-dependent associative memory $\boldsymbol{M}_{OL,ER}$ given by

$$\begin{aligned}
\boldsymbol{M}_{OL,ER} = \;& \eta\,\boldsymbol{r}_{left}\,({\boldsymbol{s}_{low}}^T \otimes {\boldsymbol{p}_{odd}}^T) + \\
& \eta'\,\boldsymbol{r}_{right}\,({\boldsymbol{s}_{low}}^T \otimes {\boldsymbol{p}_{even}}^T) + \\
& \eta'\,\boldsymbol{r}_{left}\,({\boldsymbol{s}_{high}}^T \otimes {\boldsymbol{p}_{odd}}^T) + \\
& \eta\,\boldsymbol{r}_{right}\,({\boldsymbol{s}_{high}}^T \otimes {\boldsymbol{p}_{even}}^T)
\end{aligned} \tag{3.7}$$

In either case, each parity vector is only associated with one response. However, the difference in scaling rates leads to stronger associations between low numbers and the left response, and high numbers and the right response. This model can capture the SNARC effect of a Westerner; an inversion of the inequality between scaling rates would lead to the reverse SNARC reported in Palestinians (Shaki et al., 2009). Moreover, the ratio between scaling rates determines the strength of the mapping of numbers onto space, so that Israeli participants are expected to have two similar scaling rates $\eta/\eta' \approx 1$. What is more, the associations depend on the numerical interval of the experiment, so that these context-dependent associative memories can account for the relative nature of the SNARC effect.

For example, under the *even-left* and *odd-right* condition, the output of the matrix when presented with the number 1 is

$$M_{EL,OR}(\boldsymbol{s}_1 \otimes \boldsymbol{p}_{odd}) = \left[\eta'\langle\boldsymbol{s}_{low},\boldsymbol{s}_1\rangle + \eta\langle\boldsymbol{s}_{high},\boldsymbol{s}_1\rangle\right]\boldsymbol{r}_{right} \tag{3.8}$$
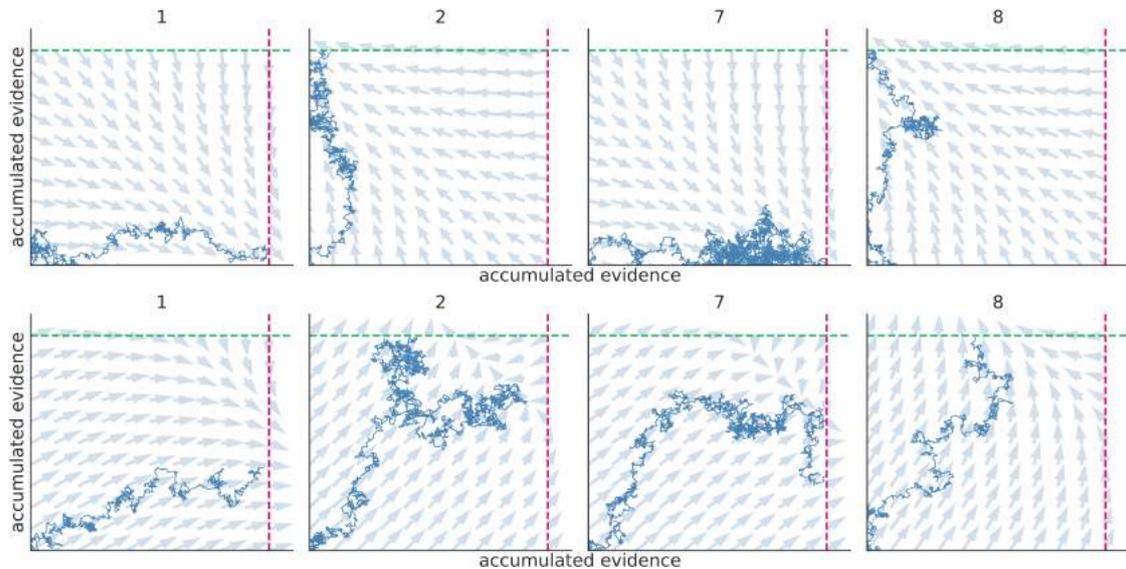
Figure 3.18: Dynamics of simulations of the SNARC effect in state space, for numbers 1, 2, 7 and 8, in an *odd-left even-right* condition. Top: orthogonal parity vectors. Bottom: non-orthogonal parity vectors. Accumulation in the latter case is more subject to noise.

whereas the same stimulus under the *odd-left* and *even-right* condition has the output

$$M_{OL,ER}(\boldsymbol{s}_1 \otimes \boldsymbol{p}_{odd}) = \left[\eta\langle\boldsymbol{s}_{low}, \boldsymbol{s}_1\rangle + \eta'\langle\boldsymbol{s}_{high}, \boldsymbol{s}_1\rangle\right]\boldsymbol{r}_{left} \tag{3.9}$$

For a conventional mapping of the MNL to a left-to-right as-if SNL, $\boldsymbol{s}_1$ is closer to $\boldsymbol{s}_{low}$ than to $\boldsymbol{s}_{high}$, so the output of the matrix in the latter case is larger than that of the former, which is consistent with faster left responses with small numbers. The same is true for the right responses and larger numbers, as well as the reversal of the SNARC effect when the ratio of scaling rates is inverted.

Next, I verify all these assertions with numerical simulations. Details about the code and parameters are included in Appendix B. Fig. 3.18 shows the accumulation in state space for different numbers.

A plot of the difference in response times for the congruent and incongruent conditions for each number is shown in Fig. 3.19. This is the MLCA model version of the original SNARC effect. The negative slope indicates faster left responses for smaller numbers and faster right responses for the larger numbers. A comparison of Fig. 3.19 with the results obtained by Dehaene et al. (1993) (Fig. 3.10) reveals two key differences. First, in the experimental results of Dehaene et al., response times appear to be grouped in pairs, with similar response time differences for 2 and 3, 4 and 5, and so on. Other versions of the same experiment do not present this behavior (e.g. Gevers et al., 2006). A second difference between the model and the experimental results is the value at which both responses are equal, i.e. the x-intercept. In the original report of the SNARC effect, the x-intercept is consistently between 2 and 4; similar results are present in some other SNARC experiments (e.g. Gevers et al., 2006) but not all (Shaki et al., 2009). I could capture both differences by modifying the numerical magnitude representations; instead, the goal here is to reproduce the SNARC effect qualitatively while knowing that, with experimental data at hand, the model is flexible enough to be fitted accordingly.

A second experiment from Dehaene et al. (1993) that I reproduce is the relative SNARC effect, that is, the fact that the difference in response times is relative to the numerical interval of the experiment (Fig. 3.10). The results are shown in Fig. 3.20.
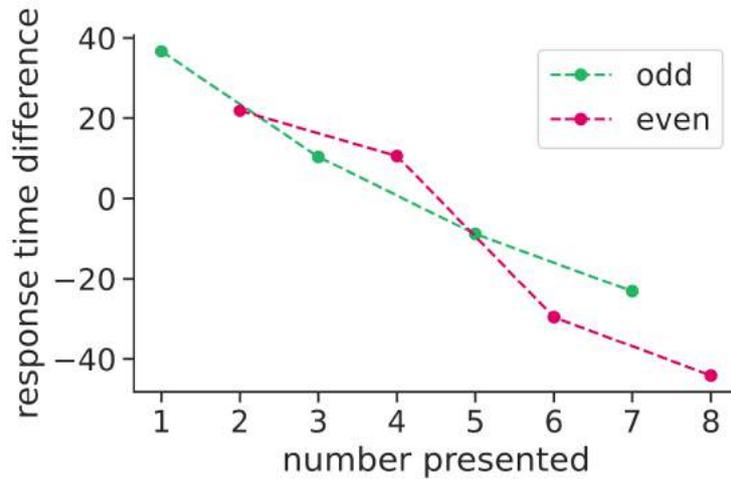
Figure 3.19: Simulation of the SNARC effect. The difference in response times between the left and the right responses is positive for low numbers and negative for high numbers.

Here, a conventional SNARC effect is observed for the numerical intervals $[1, 5]$ and $[4, 8]$. Moreover, the response times for numbers 4 and 5, which are present in both of these two sets of simulations, are different: left responses are faster for the interval $[4, 8]$ than for the interval $[1, 5]$, thus showing that the relative SNARC effect can be reproduced by mapping the relevant portion of the MNL to an as-if SNL defined by the response buttons.

A distributional analysis of the response times of the SNARC reveals a reduced effect for fast responses and a larger effect for slow responses (Gevers et al., 2006). I reproduce this experimental finding by performing a set of numerical simulations from the original SNARC and dividing the response times for each number into three categories: fast, medium and slow response times. These results are shown in Fig. 3.21. The slopes of the response times differences as a function of number increase threefold from the fast to the slow set. If one
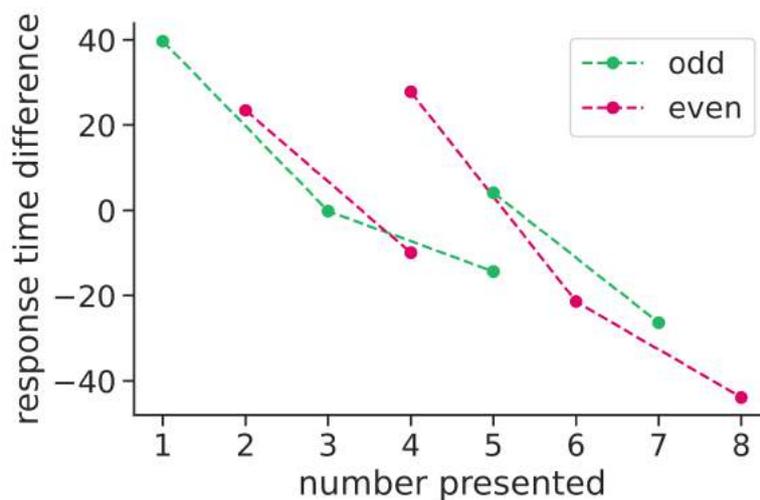
Figure 3.20: Simulation of the relative SNARC effect. Two different sets of simulations are performed, with numerical intervals $[1, 5]$ and $[4, 8]$. A conventional SNARC is found in each of these intervals.

considers this slope as indicative of the strength of the SNARC effect (Dehaene et al., 1993; Gevers et al., 2006), then these simulations reveal that the effect increases with response times. These results are equivalent to positive delta plots.

Finally, I reproduce the effect of reading direction by changing the ratio of scaling rates between the congruent and incongruent stored associations. Up to this point, I referred to the original SNARC effect as the phenomenon reported from Westerner participants with faster left responses for smaller numbers. However, Palestinians, who read text and
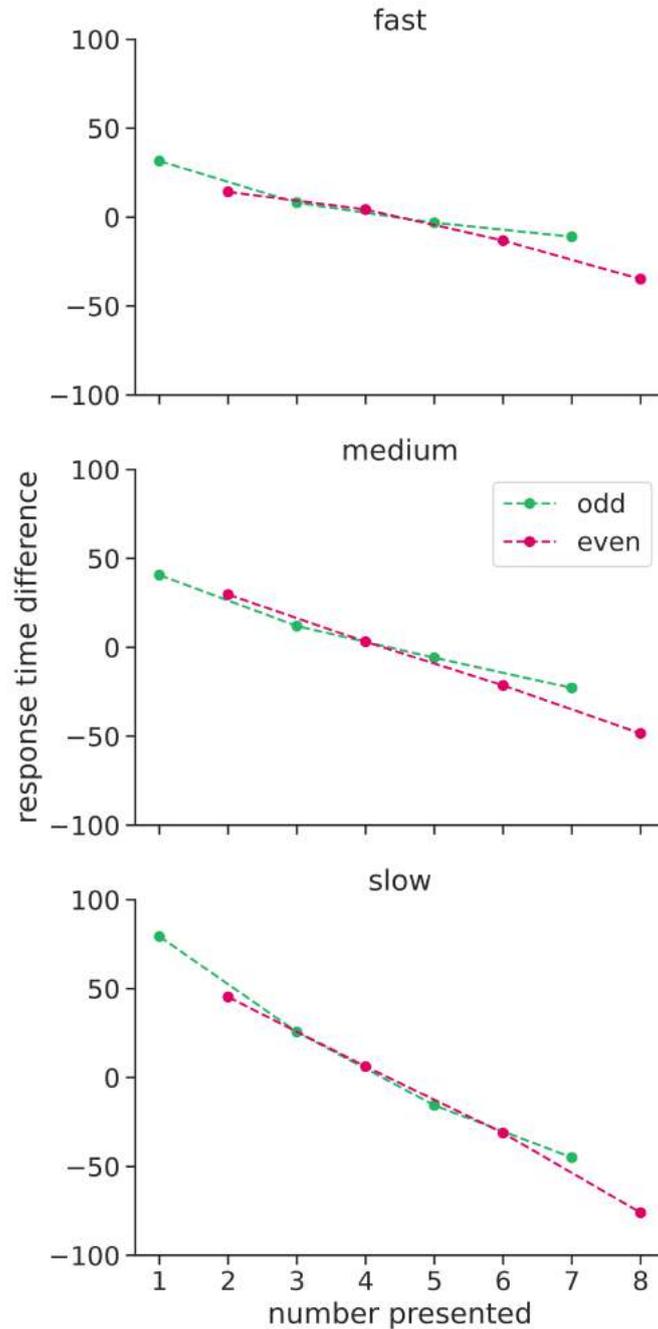
Figure 3.21: Distributional analysis of SNARC response times. All trials for each number are split into three categories according to their response times. A conventional SNARC is observed in each category, and the slope of the effect increases with response time.

numbers from right to left, present a reversed SNARC effect, whereas Israelis, who read text from right to left but numbers from left to right, present no significant SNARC (Shaki et al., 2009). These differences are modeled by making $\eta > \eta'$ for Westerners, $\eta < \eta'$ for Palestinians, and $\eta = \eta'$ for Israelis. Fig. 3.22 shows the results for each of these conditions. As predicted, there is no slope for the model of Israelis and there is a positive slope, i.e. a reversal of the SNARC, for the model of Palestinians.

To sum up, these results show that a single-route MLCA model, where the MNL is mapped to an as-if SNL along the direction of the responses, is capable of reproducing the SNARC effect under several different conditions. Parity and number are independent
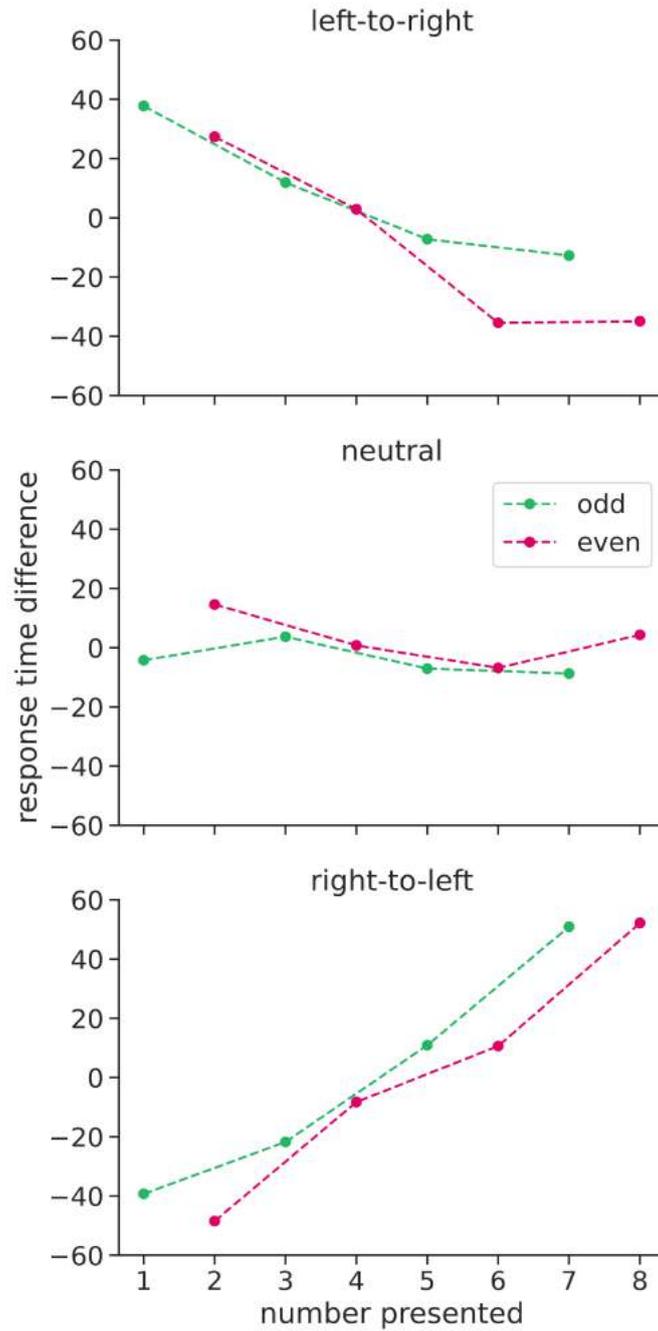
Figure 3.22: Reversal of the SNARC effect. The conventional left-to-right SNARC can be neutralized and reversed by modifying the ratio of scaling rates in (3.6) and (3.7).

attributes, in line with the distinction between visual arabic and analog magnitude representations in the triple-code model (Dehaene, 1992), which interact by means of a Kronecker product. A context-dependent associative memory stores the symmetrical number-to-space associations with respect to left and right responses. Alternatively, if fitting the model to experimental data, one can introduce an asymmetry, although this also entails an increase in the number of parameters. Previous models successfully reproduce some of the same results presented here by means of multi-layer architectures with intermediate layers (e.g. Gevers et al., 2006). In my opinion, the use of context-dependent associative memories rather than multi-layer architectures gives a better control over the associations intended to model mappings between mental representations.

3.6. Discussion

Numbers have mental representations along a one-dimensional, analog, compressed scale (Dehaene, 2003) known as the MNL. These representations are automatically activated when comparing two numbers (Restle, 1970), estimating the position of a number on a line (Siegler and Opfer, 2003), and performing basic arithmetic operations (McCrink et al., 2007), both for symbolic and non-symbolic magnitudes (Dehaene and Cohen, 1995). These number representations are likely based on a fronto-parietal network (Arsalidou and Taylor, 2011), with the IPS (Piazza et al., 2007), and in particular area VIP in monkeys (Nieder and Miller, 2004), coding for numbers irrespective of modality. Neurons tasked with identifying quantities may have emerged spontaneously from object identification in the visual system (Nasr et al., 2019).

These automatically activated number representations give rise to SRC effects with physically located responses, indicating a possible association between numerical magnitude and space (Dehaene et al., 1993; Fischer et al., 2003; Gevers et al., 2006; De Hevia et al., 2017; Rugani et al., 2015; Zorzi et al., 2002). Different theoretical accounts of these interactions include the overlap of neural codes in the IPS (Hubbard et al., 2009; Hubbard et al., 2005), long-term associations from experience and education (Shaki et al., 2009), and short-term mappings governed by working memory (Bächtold et al., 1998; Shaki and Fischer, 2008; Gevers et al., 2010; van Dijck and Fias, 2011).

The work presented in this Chapter aims at showing that these three accounts are not mutually exclusive. In fact, the theoretical model proposed in Section 3.5.2 suggests that the neural, cultural and working memory mappings can be seen as three levels of spatial-numerical associations overlaid and interacting by means of attention.

I perform numerical simulations with an MLCA model in which a number is coded by two attributes: its magnitude and its parity. These two attributes are combined by means of a Kronecker product, and the corresponding associations are stored in a context-dependent associative memory. This MLCA model successfully captures the behavior of subjects during a parity judgment task, i.e. the SNARC effect. By modifying parameters and associations, it is possible to reproduce different results from SNARC tasks, such as the relativity of response times with numerical intervals and the reversal of the SNARC, which can be achieved by short-term associations or by a long-term influence of reading direction.

A comment on functional isomorphism

Upon discussing their results from the Mundurucú (Section 3.3.2), Dehaene et al. state: "Overall, these results reveal both universal and culture-dependent facets of the sense of number. After a minimal instruction period, even members of a remote culture with reduced vocabulary and education readily understand that number can be mapped onto a spatial scale" (Dehaene et al., 2008, p. 1218). The need for an instruction period indicates that the number-to-space mappings are straightforward but not spontaneously emerging.

A critique of the Mundurucú findings by Cantlon et al. (2009) argues that the conclusions drawn by Dehaene et al. (2008) are misleading. Cantlon et al. compare the Mundurucú experiment to a study by Rattermann and Gentner (1998), who find that young children can associate the concepts *daddy*, *mommy*, and *baby* to *large*, *medium*, and *small* flower pots. This does not, according to Cantlon et al., mean that the children's underlying representations of a family are mapped to three different-sized flower pots. This comment by Cantlon et al. (2009) is relevant in the context of the present thesis. Number-space mappings, and particularly SNLs, do not necessarily indicate that the underlying representations of numbers and space are permanently associated, but merely that, for each specific task at hand, numbers can be mapped onto space.

In this sense, I believe a functional isomorphism between numbers and space, as suggested by Zorzi et al. (2002), may be a more parsimonious account. In the words of Hillary Putnam: "The concept which is key to unraveling the mysteries in the philosophy of mind, I think, is the concept of functional isomorphism. Two systems are functionally isomorphic if there is a correspondence between the states of one and the states of the other that preserves functional relations." (Putnam, 1975, p.291-292).

Without adhering to a functionalist view of the mind, one can nevertheless identify the advantages of focusing on associations between cognitive functions rather than uncritically restraining our theories to potentially permanent mappings such as shared neural representations. This is, at least, until technological advancements in neuroimaging and electrophysiological experiments allow us to directly uncover the simultaneous real-time activity of entire populations of neurons.

Three number lines

The way in which I approach the associations between numbers and space in this thesis is by drawing distinctions between three different number lines: the MNL, an analog magnitude representation in the mind; the SNL, an explicit number representation along existing physical lines; and the as-if SNL, an implicit number representation along non-existing physical lines. My account of the spatial-numerical associations revealed by the SNARC effect consists of a mapping of the MNL onto as-if SNLs in the same way as it is mapped onto SNLs. When numerical magnitude is task-irrelevant, as is the case of the SNARC effect, this mapping affects response times with hemispatial asymmetry.

Previous theoretical accounts suggest a multi-dimensional representation for numbers, with magnitude and space being the two dimensions involved in the SNARC effect. In my view, these accounts undermine the role of working memory. The use of multiple number lines, with mappings between them, explains spatial-numerical associations in a more comprehensive way.

Further insights into the representation of numbers

Previous works with sequential sampling models such as the LCA and the DDM (see Chapter 1) within numerical cognition and related fields suggest these can be successfully implemented to account for number recognition (Schmiedek et al., 2007), semantic congruity (Pirrone et al., 2017), hysteresis effects (Odic et al., 2014), and cognitive development with age (Ratcliff and McKoon, 2008). The models developed in this Section, i.e. the LCA model for number comparisons and the MLCA model for parity judgments, can be readily adapted to reproduce a myriad of existing number-based decisions. Here, my aim was to capture spatial-numerical associations, and so the focus was mainly set on the SNARC effect. Other effects, which may or may not include a spatial component, can likely also be modeled by making minimal changes to the MLCA model.

Conclusions

This thesis was motivated by the idea that the SNARC could be seen as a contextual influence of the task-irrelevant numerical magnitude over the task-relevant parity, thus producing the characteristic time differences of the effect. This hypothesis led me on a path to develop both a computational model and a theoretical account of number-space mappings.

For the computational model, I selected the LCA model of Usher and McClelland (2001) as a starting point, and introduced a novel vector-matrix formulation as well as the use of context-dependent associative memories (Mizraji, 1989). These two innovations allowed me to delve into the representational domain, with vectors identifying the mental representations of different attributes of stimuli. The multi-attribute extension of the LCA, which I called the MLCA, was successfully implemented to reproduce behavioral results from a multi-attribute RDK task with monkeys (Mante et al., 2013). A dual-route version of the same model was also capable of capturing the Simon effect (Simon and Small, 1969). Finally, the MLCA model was adapted to reproduce the numerical distance effect (Hinrichs et al., 1981) and several different versions of the SNARC effect (Dehaene et al., 1993), with numerical magnitude and parity being two independent attributes (Dehaene, 1992) which interact by means of a Kronecker product.

Concurrently, a theoretical account was given to explain seemingly contradictory experimental results about spatial-numerical associations. Previously, three different origins had been suggested: a neural overlap (Hubbard et al., 2005), long-term cultural effects (Shaki et al., 2009), and short-term mappings in working memory (Fias and van Dijck, 2016). As is often the case, the correct answer likely lies somewhere in the middle. I suggest that each of these three different origins constitutes a level in a triple overlay. Humans and non-human animals may be born with innate spatial-numerical associations. As each individual develops and interacts with the world, and particularly by means of formal education and the experience of reading and writing in a fixed direction, long-term mappings between numbers and space are strengthened, quasi-permanently. However, these associations can be reversed, either by being immersed in a culture with a different reading direction (Dehaene et al., 1993) or temporarily with ingenious experiments which involve working memory (van Dijck and Fias, 2011). All three levels of this theoretical account contribute to the SNARC effect and other spatial-numerical associations, while attentional resources are required for higher levels to be properly employed.

To sum up, the model developed in this thesis constitutes a representational extension of an existing computational model, which can adequately approximate the biological mechanism for multi-attribute evidence integration, and is backed by a comprehensive theoretical account of what these attributes are and how they may interact.

Although the MLCA was specifically developed to reproduce number-space mappings, it can be readily adapted to capture other effects. This is done at the representational level, by modifying the vectors which capture the mesoscopic levels of activity associated with different classes of stimuli. This prompts the question of its generalizability. More often than not, computational models in cognitive science are restricted to a single

phenomenon or effect, so that little information is obtained about the way in which the brain processes information in general. I do not wish to claim that the work presented here can be immediately interpreted in the context of other cognitive functions. Nevertheless, I do believe that this research can be seen as a contribution towards the development of a general-purpose model of cognition.

This work can be extended in several directions. A first possibility, in the manner of traditional computational studies in perceptual decision-making, would be to focus on the optimization of the integration algorithm and on finding the best fits for the different parameters. This requires experimental data from established experimental paradigms with well-accepted results. This approach would lead to an improvement of the computational nature of the model by matching its predictions to human and non-human animal behavior. Because the MLCA is based on the LCA, one should expect the results obtained with the original model to extend naturally to this novel proposal. However, a quantitative validation of the MLCA would be of great interest.

A second line of work would be to design and implement a series of experiments to verify the predictions of the MLCA model for which no behavioral evidence is available as of yet. Most notably, in Chapter 1 I briefly mentioned that the way in which multiple attributes are mathematically combined in the context-dependent associative memories leads to an increase in response times with increasing number of alternatives. Intuitively this seems reasonable. A verification is nonetheless much needed. Should any differences between these predictions and the results of such an experiment be revealed, an adaptation of the MLCA model would be necessary. This approach would lead to an improvement of the representational nature of the model.

Finally, a topic which was discussed but not included in the model is the process of learning. Associative memory models can be built from one-shot Hebbian learning, as was done here, or by a number of other methods designed both to optimize storage and to approximate natural learning. In this sense, it would be of value to explore different learning methods and whether they can reproduce learning across two different time scales: changes through ontogenetic development, for example changes to the structure of the MNL, and also performance improvements from the beginning to the end of an experiment. This approach would lead to a validation of MLCA as a model of human and non-human animal cognition.

In closing, I would like to share my views on the role of theoretical and mathematical models in cognitive science. Over the last century or so, an ever-growing corpus of behavioral and neural experiments has started to reveal the workings of the mind. I say started because, even if we now know much more than just a few decades ago, and even though the methodologies and technologies are constantly improving, there is much that cannot yet be measured experimentally. The gap between the neural substrate and the resulting behavior is still too wide, and computational models are essential to connect these levels. Theories, however, often follow one step behind. And most of these are of a descriptive or explanatory nature. As cognitive science matures as a discipline of its own, independent of psychology and neuroscience, I forecast predictive theories and mathematical models will be of central importance.

In the preface to his book *Design for a brain*, W. Ross Ashby writes: "Having experienced the confusion that tends to arise whenever we try to relate cerebral mechanisms to observed behaviour, I made it my aim to accept nothing that could not be stated in mathematical form, for only in this language can one be sure, during one's progress, that one is not unconsciously changing the meaning of terms, or adding assumptions, or otherwise drifting towards confusion." (Ashby, 1952, p. 5). I make these words my own.

Appendix A

From perceptions to representations

Throughout this thesis, external stimuli are considered to be equivalent to their mental representations. This is an assumption of perfect discrimination, which the animal brain is known to be flawed at. Nevertheless, the assumption holds for the tasks studied. Prior to forming the mental representations of stimuli, they are perceived by sensory receptors (e.g. the eye) and activate the corresponding primary sensory cortex (e.g. area V1) and successive areas of the visual system (e.g. area MT). Perfect discrimination may be possible with a set of very different stimuli.

The conversion from a stimulus perception to its representation, as performed for example by direction-selective neurons in the visual cortex, can be modeled with a class of associative memory called an autoassociative memory. Unlike the heteroassociative memories introduced in Section 1.2.1, autoassociative memories store associations between a vector and itself. When a noisy version of that vector is used as input, the memory outputs the originally stored vector, provided other ones have low interference.

Let $\boldsymbol{f_1}$ and $\boldsymbol{f_2}$ be two stimuli, for example leftwards and rightwards moving dots in a RDK. An autoassociative memory $\boldsymbol{M_A}$ can be constructed as

$$\boldsymbol{M_A} = \boldsymbol{f_1}\boldsymbol{f_1^T} + \boldsymbol{f_2}\boldsymbol{f_2^T} \tag{A.1}$$

Now let c be an experimental coherence, for example the motion coherence for a RDK experiment, with $0 \leq c \leq 1$. Consider the stimulus

$$\boldsymbol{f} = c\boldsymbol{f_1} + (1-c)\left[\frac{\boldsymbol{f_1}}{2} + \frac{\boldsymbol{f_2}}{2}\right] = \frac{(1+c)}{2}\boldsymbol{f_1} + \frac{(1-c)}{2}\boldsymbol{f_2} \tag{A.2}$$

where a fraction c of the dots move coherently as $\boldsymbol{f_1}$ and the remaining $(1-c)$ move randomly, half of them in each direction. Using $\boldsymbol{f}$ as an input to the autoassociative memory gives

$$\boldsymbol{M_A}\boldsymbol{f} = \frac{(1+c)}{2}\langle\boldsymbol{f_1},\boldsymbol{f_1}\rangle\boldsymbol{f_1} + \frac{(1-c)}{2}\langle\boldsymbol{f_2},\boldsymbol{f_2}\rangle\boldsymbol{f_2} + \frac{(1+c)}{2}\langle\boldsymbol{f_1},\boldsymbol{f_2}\rangle\boldsymbol{f_1} + \frac{(1-c)}{2}\langle\boldsymbol{f_2},\boldsymbol{f_1}\rangle\boldsymbol{f_2}$$

$$\tag{A.3}$$

Assuming the stored vectors are orthonormal, i.e. $\langle\boldsymbol{f_i},\boldsymbol{f_j}\rangle = \delta_{ij}$, the output of the autoassociative memory is given by

$$\boldsymbol{M_A}\boldsymbol{f} = \frac{(1+c)}{2}\boldsymbol{f_1} + \frac{(1-c)}{2}\boldsymbol{f_2} = \boldsymbol{f} \tag{A.4}$$

so that the output of the matrix coincides with the input. In other words, the representation for a noisy stimulus is the same as the noisy stimulus itself, provided the stored stimuli

are orthogonal, i.e. very different, and the noise is formed by a linear combination of those stimuli.

Similar results can be obtained with context-dependent autoassociative memories that store the associations of vectors formed by Kronecker products. Such a system can form the representation of some or all of the attributes of a stimulus.

An autoassociative memory is a very elementary model of how a representation can be extracted from a perception in a single step. Alternatively, recurrent dynamics might be a more suitable model for complex tasks. The BSB (Anderson et al., 1977) briefly explained in Section 2.4.1 and Hopfield networks are examples of recurrent dynamical models which use autoassociative memories to update the vector values, a process that continues until convergence, ideally to one of the stored vectors. Visual perception has both bottom-up and top-down processing, so recurrent dynamics should not be immediately discarded.

Nonetheless, a one-step autoassociative memory is sufficient to show that stimuli representations can be formed from noisy stimuli, and that using the stimulus vector as a representation of the stimulus itself is not unreasonable.

Appendix B

Details of numerical simulations

B.1. Integration algorithm

All numerical simulations were performed using Python version 3.8.5. The parameters of each simulation are detailed in the following Sections; the integration algorithm was common in all cases and is presented in pseudocode next.

```
0.  BEGIN
1.  SET PARAMETERS
2.  SET STORED VECTORS
3.  GET ASSOCIATIVE MEMORY FROM PARAMETERS AND STORED VECTORS
4.  SET PRESENTED VECTOR
5.  SET TOTAL ACCUMULATION TO ZERO
6.  SET INTEGRATION TIME TO ZERO
7.  WHILE (INTEGRATION TIME IS LESS THAN TOTAL TIME):
8.    GET PRODUCT OF ASSOCIATIVE MEMORY AND PRESENTED VECTOR
9.    GET RANDOM NOISE
10.   GET ACCUMULATION DIFFERENTIAL
11.   GET SUM OF TOTAL ACCUMULATION AND ACCUMULATION DIFFERENTIAL
12.   IF (SUM IS LARGER THAN THRESHOLD):
13.     RETURN INTEGRATION TIME, TOTAL ACCUMULATION
14.     END
15.   ELSE IF (SUM IS NEGATIVE):
16.     SET TOTAL ACCUMULATION TO ZERO
17.   ELSE:
18.     SET TOTAL ACCUMULATION TO SUM
19. RETURN INTEGRATION TIME, TOTAL ACCUMULATION
20. END
```

B.2.   Multi-attribute, multi-alternative decisions

In Section 1.3.2, numerical simulations were performed for 2, 4, 8, 16, and 32 alternatives with a single attribute (Fig. 1.7). For each, 100 trials were simulated with the integration parameters shown in table B.1 and a presented stimulus given by a correct response for the first accumulator, that is

$$s = [1, 0, ..., 0]^T = e_1 \tag{B.1}$$

where e_1 is the standard basis vector associated with the first accumulator. The input towards the remaining accumulators comes exclusively from the noisy integration. Increasing the number of alternatives entails an increase in the inhibition towards the correct accumulator, and therefore a slower accumulation.

Also in Section 1.3.2, numerical simulations were performed for a two-alternative choice with 1, 2, 4, 6, 8, and 12 attributes for two coherences of 10% and 50% (Fig. 1.8). 1000 trials were simulated with random normalized stimuli vectors and the parameters shown in table B.1. Given n attributes, the associative memory $M_{(2 \times n)}$ is defined to store the associations between the first attribute, s_1, and the response vectors, while the remaining $n - 1$ attributes are task-irrelevant. To do so, the response vectors and the first attribute vectors are the standard basis vectors $\{e_1, e_2\}$, and the remaining attributes are normalized all-ones vectors. This is done by making (1.16) with

$$\begin{aligned}
r^{(1)} &= [1, 0]^T = e_1, & r^{(2)} &= [0, 1]^T = e_2 \\
s_1^{(1)} &= [1, 0]^T = e_1, & s_2^{(2)} &= [0, 1]^T = e_2 \\
s_i^{(1)} &= [1/m, 1/m]^T, & s_i^{(2)} &= [1/m, 1/m]^T
\end{aligned} \tag{B.2}$$

After M is established, for each different simulation, the n attribute vectors are sampled as 2-dimensional random vectors each and normalized. Next, the first attribute vector s_1 is updated according to the relevant coherence level (either $c = 10\%$ or $c = 50\%$) by adding $c\, e_{1,2}$ to it, where $e_{1,2}$ is the standard basis vector corresponding to the correct response.

Table B.1: MLCA parameters used to simulate multi-attribute and multi-alternative decisions in Section 1.3.2.

Parameter	Name	Value
α	Leakage	0.02
β	Lateral inhibition	0.02
dt	Integration step	0.1
τ	Time scale	1
σ	Gaussian noise variance	0.25
b	Threshold boundary	10
η	Scaling factor	1

B.3. Multi-attribute RDK

In Section 1.4, the numerical simulations for monkey A (Mante et al., 2013) were performed with the parameters shown in table B.2 and the context-dependent associative memory $\mathcal{M}$ defined in (1.28) with

$$
\begin{aligned}
\boldsymbol{r}_{left} &= [1, 0]^T \\
\boldsymbol{r}_{right} &= [0, 1]^T \\
\boldsymbol{c}_{motion} &= [1, 0]^T \\
\boldsymbol{c}_{color} &= [0, 1]^T \\
\boldsymbol{p}_{left} &= [\sqrt{1 - 1/9}, \sqrt{1/9}]^T \\
\boldsymbol{p}_{right} &= [\sqrt{1/9}, \sqrt{1 - 1/9}]^T \\
\boldsymbol{q}_{green} &= [\sqrt{1 - 1/6}, \sqrt{1/6}]^T \\
\boldsymbol{q}_{red} &= [\sqrt{1/6}, \sqrt{1 - 1/6}]^T \\
\eta &= 1 \\
\eta' &= 0.8
\end{aligned}
\tag{B.3}
$$

and for monkey F the context-dependent associative memory defined by the same vectors except for the context cues, which included a small overlap to capture this monkey's worse results:

$$
\begin{aligned}
\boldsymbol{c}_{motion} &= [1, 0]^T \\
\boldsymbol{c}_{color} &= [\sqrt{1/16}, \sqrt{1 - 1/16}]^T
\end{aligned}
\tag{B.4}
$$

The numerical simulations following the dynamics of (1.27) were performed with the same context cues used for the memories, while the stimuli varied according to Fig. 1.9 across 11 coherence levels:

$$
\begin{aligned}
\boldsymbol{p}_{motion} &= [\sqrt{1 - i/10}, \sqrt{i/10}]^T, \quad i \in \{0, 1, ..., 10\} \\
\boldsymbol{c}_{color} &= [\sqrt{1 - i/10}, \sqrt{i/10}]^T, \quad i \in \{0, 1, ..., 10\}
\end{aligned}
\tag{B.5}
$$

Table B.2: MLCA parameters used to simulate the monkeys' decisions during the experiment performed by Mante et al. (2013), as described in Section 1.4.

Parameter	Name	Value
α	Leakage	0.1
β	Lateral inhibition	0.1
dt	Integration step	0.1
τ	Time normalization	1
σ	Gaussian noise variance	0.25
b	Threshold boundary	10

B.4. Simon effect

The numerical simulations of the Simon effect were performed with the dual-route MLCA model presented in Section 2.4.2 with the parameters shown in table B.3. The vectors used for the discrimination of a square and a circle (see Fig. 2.4) in (2.6) and (2.8) were given by

$$
\begin{aligned}
\boldsymbol{r}_{left} &= [1, 0]^T \\
\boldsymbol{r}_{right} &= [0, 1]^T \\
\boldsymbol{f}_{square} &= [1, 0]^T \\
\boldsymbol{f}_{circle} &= [0, 1]^T \\
\boldsymbol{s}_{left} &= [1, 0]^T \\
\boldsymbol{s}_{right} &= [0, 1]^T
\end{aligned}
\tag{B.6}
$$

For the controlled process the scaling rate was fixed at $\eta_c = 0.02$; for the automatic process two different situations were considered. An early time-to-peak, with $\eta = 0.25$, $\eta' = 0.125$ which lasted for $T = 20$ units of time, and a late time-to-peak, with $\eta = 0.02$, $\eta' = 0.01$ which lasted for $T = 500$ units of time. 1000 trials were simulated for each of Fig. 2.13 and Fig. 2.14.

Table B.3: MLCA parameters used to simulate the Simon effect, as described in Section 2.4.2.

Parameter	Name	Value
α	Leakage	0.005
β	Lateral inhibition	0.005
dt	Integration step	0.1
τ	Time normalization	1
σ	Gaussian noise variance	0.2
b	Threshold boundary	10

B.5. Numerical distance effect

In Section 3.5.1, numerical simulations of an LCA model of the distance effect according to (3.1) were performed with the parameters shown in table B.4 and the following vectors:

$$
\begin{aligned}
\boldsymbol{r}_{lower} &= [1,0]^T \\
\boldsymbol{r}_{higher} &= [0,1]^T \\
\boldsymbol{s}_{lower} &= [1,0]^T \\
\boldsymbol{s}_{higher} &= [0,1]^T
\end{aligned}
\tag{B.7}
$$

First, Fig. 3.15 and Fig. 3.16 were obtained with a linear mapping of numbers, with $\boldsymbol{s}_{lower}$ corresponding to the number 11, $\boldsymbol{s}_{higher}$ to the number 99, and any other number n in between attributed a linear combination of the two, i.e. $\boldsymbol{s} = \gamma\boldsymbol{s}_{lower} + (1-\gamma)\boldsymbol{s}_{higher}$, where $\gamma = (n-11)/(99-11)$. Three trials were simulated for each number between 11 and 53 and between 57 and 99.

The simulations of Fig. 3.17 were done with a logarithmic mapping of numbers, with $\boldsymbol{s}_{lower}$ corresponding to the number 11, $\boldsymbol{s}_{higher}$ to its symmetric number on a logarithmic scale, i.e. 275, and any other number n in between attributed a linear combination of the two, i.e. $\boldsymbol{s} = \gamma\boldsymbol{s}_{lower} + (1-\gamma)\boldsymbol{s}_{higher}$, where $\gamma = (\log(n) - \log(11))/(\log(275) - \log(11))$. Again, three trials were simulated for each number between 11 and 53 and between 57 and 99.

Table B.4: MLCA parameters used to simulate the numerical distance effect, as described in Section 3.5.1.

Parameter	Name	Value
α	Leakage	0.005
β	Lateral inhibition	0.005
dt	Integration step	0.1
τ	Time normalization	1
σ	Gaussian noise variance	0.05
b	Threshold boundary	10

B.6. SNARC effect

The simulations of the MLCA model of the SNARC presented in Section 3.5.3 were performed with the parameters shown in table B.5. The context-dependent associative memories (3.6) and (3.7) were defined with

$$
\begin{aligned}
\boldsymbol{r}_{left} &= [1,0]^T \\
\boldsymbol{r}_{right} &= [0,1]^T \\
\boldsymbol{s}_{low} &= [1,0]^T \\
\boldsymbol{s}_{high} &= [0,1]^T
\end{aligned}
\tag{B.8}
$$

and the parity vectors for the *even-left odd-right* condition were

$$
\begin{aligned}
\boldsymbol{p}_{even} &= [1,0]^T \\
\boldsymbol{p}_{odd} &= [0,1]^T
\end{aligned}
\tag{B.9}
$$

whereas for the *odd-left even-right* condition were

$$
\begin{aligned}
\boldsymbol{p}_{odd} &= [1,0]^T \\
\boldsymbol{p}_{even} &= [0,1]^T
\end{aligned}
\tag{B.10}
$$

For the original SNARC, 300 trials were simulated for each number between 1 and 8, with 1 corresponding to $\boldsymbol{s}_{low}$ and 8 to $\boldsymbol{s}_{high}$, and scaling rates $\eta = 1$ and $\eta' = 0.8$. Those 300 simulations were split into fast, medium and slow response times for the distributional analysis. For the relative SNARC, 100 trials were simulated for numbers between 1 and 5, with 1 corresponding to $\boldsymbol{s}_{low}$ and 5 to $\boldsymbol{s}_{high}$, and another 100 trials were simulated for numbers between 4 and 8, with 4 corresponding to $\boldsymbol{s}_{low}$ and 8 to $\boldsymbol{s}_{high}$. The reversal of the SNARC is modeled by changing the scaling rates between the left-to-right ($\eta = 1$, $\eta' = 0.8$), neutral ($\eta = 0.9, \eta' = 0.9$), and right-to-left ($\eta = 0.8, \eta' = 1$) conditions.

Table B.5: MLCA parameters used to simulate the SNARC effect, as described in Section 3.5.3.

Parameter	Name	Value
α	Leakage	0.005
β	Lateral inhibition	0.005
dt	Integration step	0.1
τ	Time normalization	1
σ	Gaussian noise variance	0.25
b	Threshold boundary	10

Bibliography

Agrillo, C., Piffer, L., Bisazza, A., & Butterworth, B. (2012). Evidence for two numerical systems that are similar in humans and guppies. *PloS one*, *7*(2), e31923.

Andersen, R. A., Snyder, L. H., Bradley, D. C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual review of neuroscience*, *20*(1), 303–330.

Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical biosciences*, *14*(3-4), 197–220.

Anderson, J. A. (1995). *An introduction to neural networks*. Cambridge, MA: MIT Press.

Anderson, J. A. (2003). Associative networks. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (2nd ed., pp. 117–122). Cambridge, MA: MIT Press.

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological review*, *84*(5), 413.

Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459.

Arsalidou, M., & Taylor, M. J. (2011). Is 2+ 2= 4? meta-analyses of brain areas needed for numbers and calculations. *Neuroimage*, *54*(3), 2382–2393.

Ashby, W. R. (1952). *Design for a brain: The origin of adaptive behaviour*. Wiley.

Attout, L., Fias, W., Salmon, E., & Majerus, S. (2014). Common neural substrates for ordinal representation in short-term memory, numerical and alphabetical cognition. *PloS one*, *9*(3), e92049.

Azañón, E., Longo, M. R., Soto-Faraco, S., & Haggard, P. (2010). The posterior parietal cortex remaps touch into external space. *Current biology*, *20*(14), 1304–1309.

Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational space. *Neuropsychologia*, *36*(8), 731–735.

Baker Jr, C. L., & Braddick, O. J. (1982). The basis of area and dot number effects in random dot motion perception. *Vision research*, *22*(10), 1253–1259.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *The Journal of physiology*, *193*(2), 327–342.

Barry, C., Lever, C., Hayman, R., Hartley, T., Burton, S., O'Keefe, J., Jeffery, K., & Burgess, N. (2006). The boundary vector cell model of place cell firing and spatial memory. *Reviews in the Neurosciences*, *17*(1-2), 71.

Beck, J. M., Latham, P. E., & Pouget, A. (2011). Marginalization in neural circuits with divisive normalization. *Journal of Neuroscience*, *31*(43), 15310–15319.

beim Graben, P., & Potthast, R. (2009). Inverse problems in dynamic cognitive modeling. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *19*(1), 015103.

Beran, M. J., French, K., Smith, T. R., & Parrish, A. E. (2019). Limited evidence of number–space mapping in rhesus monkeys (macaca mulatta) and capuchin monkeys (sapajus apella). *Journal of Comparative Psychology*, *133*(3), 281.

Birman, D., & Gardner, J. L. (2019). A flexible readout mechanism of human sensory representations. *Nature communications*, *10*(1), 1–13.

Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*(1), 129–133.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review*, *113*(4), 700.

Bogacz, R., Usher, M., Zhang, J., & McClelland, J. L. (2007). Extending a biologically inspired model of choice: Multi-alternatives, nonlinearity and value-based multidimensional choice. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1655–1670.

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in neurosciences*, *33*(1), 10–16.

Braddick, O. (1974). A short-range process in apparent motion. *Vision research*, *14*(7), 519–527.

Brown, S., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology*, *53*(6), 453–462.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple scalability, fixed-sample, and sequential-sampling models. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(3), 538.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, *100*(3), 432.

Cantlon, J. F., & Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biol*, *5*(12), e328.

Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on" log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures". *science*, *323*(5910), 38–38.

Cheng, K. (1986). A purely geometric module in the rat's spatial representation. *Cognition*, *23*(2), 149–178.

Churchland, A. K., & Kiani, R. (2016). Three challenges for connecting model to mechanism in decision-making. *Current opinion in behavioral sciences*, *11*, 74–80.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature neuroscience*, *11*(6), 693–702.

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, *3*(7), 553–562.

Colby, C. L., & Goldberg, M. E. (1999). SPACE AND ATTENTION IN PARIETAL CORTEX. *Annual Review of Neuroscience*, *22*(1), 319–349.

Cooper, L. N. (1973). A possible organization of animal memory and learning. In B. Lundquist & S. Lundquist (Eds.), *Proceedings of the nobel symposium on collective properties of physical systems* (pp. 252–264). New York: Academic Press.

Cooperrider, K., Marghetis, T., & Núñez, R. (2017). Where does the ordered line come from? evidence from a culture of papua new guinea. *Psychological science*, *28*(5), 599–608.

Courtière, A., Hardouin, J., Burle, B., Vidal, F., & Hasbroucq, T. (2007). Simon effect in the rat: A new model for studying the neural bases of the dual-route architecture. *Behavioural brain research*, *179*(1), 69–75.

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413–1420.

De Hevia, M. D., Girelli, L., Addabbo, M., & Cassia, V. M. (2014). Human infants' preference for left-to-right oriented increasing numerical sequences. *PloS one*, *9*(5), e96412.

De Hevia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science*, *21*(5), 653–660.

De Hevia, M. D., Veggiotti, L., Streri, A., & Bonn, C. D. (2017). At birth, humans associate "few" with left and "many" with right. *Current Biology*, *27*(24), 3879–3884.

De Hevia, M.-D., & Spelke, E. S. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition*, *110*(2), 198–207.

De Jong, R., Liang, C.-C., & Lauber, E. (1994). Conditional and unconditional automaticity: A dual-process model of effects of spatial stimulus-response correspondence. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(4), 731.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1-2), 1–42.

Dehaene, S. (2003). The neural basis of the weber–fechner law: A logarithmic mental number line. *Trends in cognitive sciences*, *7*(4), 145–147.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of experimental psychology: General*, *122*(3), 371.

Dehaene, S., & Changeux, J.-P. (1993). Development of elementary numerical abilities: A neuronal model. *Journal of cognitive neuroscience*, *5*(4), 390–407.

Dehaene, S., & Cohen, L. (1995). Towards an anatomical and functional model of number processing. *Mathematical cognition*, *1*(1), 83–120.

Dehaene, S., & Cohen, L. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? analogical and symbolic effects in two-digit number comparison. *Journal of experimental Psychology: Human Perception and performance*, *16*(3), 626.

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or linear? distinct intuitions of the number scale in western and amazonian indigene cultures. *science*, *320*(5880), 1217–1220.

Dehaene, S., Tzourio, N., Frak, V., Raynaud, L., Cohen, L., Mehler, J., & Mazoyer, B. (1996). Cerebral activations during number multiplication and comparison: A pet study. *Neuropsychologia*, *34*(11), 1097–1106.

de Lafuente, V., Jazayeri, M., & Shadlen, M. N. (2015). Representation of accumulating evidence for a decision in two parietal areas. *Journal of Neuroscience*, *35*(10), 4306–4318.

Descartes, R. (1637). *La géométrie*. Ian Maire.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, *41*(3), 260–274.

Diederich, A. (2003). Mdft account of decision making under time pressure. *Psychonomic Bulletin & Review*, *10*(1), 157–166.

Diederich, A. (2016). A multistage attention-switching model account for payoff effects on perceptual decision tasks with manipulated processing order. *Decision*, *3*(2), 81.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388–1404.

Diederich, A., & Oswald, P. (2014). Sequential sampling model for multiattribute choice alternatives with random attention time and processing order. *Frontiers in human neuroscience*, *8*, 697.

Ditterich, J., Mazurek, M. E., & Shadlen, M. N. (2003). Microstimulation of visual cortex affects the speed of perceptual decisions. *Nature neuroscience*, *6*(8), 891–898.

Donner, T. H., Siegel, M., Fries, P., & Engel, A. K. (2009). Buildup of choice-predictive activity in human motor cortex during perceptual decision making. *Current Biology*, *19*(18), 1581–1585.

Drucker, C. B., & Brannon, E. M. (2014). Rhesus monkeys (macaca mulatta) map number onto space. *Cognition*, *132*(1), 57–67.

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1998). Ventral intraparietal area of the macaque: Congruent visual and somatic response properties. *Journal of neurophysiology*, *79*(1), 126–136.

Eimer, M. (1997). An event-related potential (erp) study of transient and sustained visual attention to color and form. *Biological Psychology*, *44*(3), 143–160.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143–149.

Fechner, G. T. (1860). *Elemente der psychophysik* (Vol. 2). Breitkopf u. Härtel.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314.

Fias, W., & van Dijck, J.-P. (2016). The temporary nature of number—space interactions. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(1), 33.

Fias, W., van Dijck, J.-P., & Gevers, W. (2011). How is number associated with space? the role of working memory. *Space, time and number in the brain*, 133–148.

Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts of attention. *Nature neuroscience*, *6*(6), 555–556.

Fitts, P. M., & Deininger, R. L. (1954). Sr compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of experimental psychology*, *48*(6), 483.

Fitts, P. M., & Seeger, C. M. (1953). Sr compatibility: Spatial characteristics of stimulus and response codes. *Journal of experimental psychology*, *46*(3), 199.

Fowler, D., & Robson, E. (1998). Square root approximations in old babylonian mathematics: Ybc 7289 in context. *Historia Mathematica*, *25*(4), 366–378.

Friston, K. J. (2011). Functional and effective connectivity: A review. *Brain connectivity*, *1*(1), 13–36.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1-2), 43–74.

Galton, F. (1880). Visualised numerals. *Nature*, *21*(533), 252–256.

Georgopoulos, A. P., Lurito, J. T., Petrides, M., Schwartz, A. B., & Massey, J. T. (1989). Mental rotation of the neuronal population vector. *Science*, *243*(4888), 234–236.

Gerstmann, J. (1940). Syndrome of finger agnosia, disorientation for right and left, agraphia and acalculia: Local diagnostic value. *Archives of Neurology & Psychiatry*, *44*(2), 398–408.

Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., & Verguts, T. (2010). Verbal-spatial and visuospatial coding of number–space interactions. *Journal of Experimental Psychology: General*, *139*(1), 180.

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the snarc effect. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(1), 32.

Gilbert, C. D., Sigman, M., & Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, *31*(5), 681–697.

Gold, J. I., & Ding, L. (2013). How mechanisms of perceptual decision-making affect the psychometric function. *Progress in neurobiology*, *103*, 98–114.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual review of neuroscience*, *30*.

Goldstone, R. L. (1998). Perceptual learning. *Annual review of psychology*, *49*(1), 585–612.

Gómez Haedo, J. C. (1937). Un vocabulario charrúa desconocido. *Boletín de Filología*, *1*(4-5), 323–350.

Gordon, P. (2004). Numerical cognition without words: Evidence from amazonia. *Science*, *306*(5695), 496–499.

Graham, A. (1981). *Kronecker products and matrix calculus with applications*. Chichester : Ellis Horwood.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*(1), 3–17.

Grieves, R. M., & Jeffery, K. J. (2017). The representation of space in the brain. *Behavioural processes*, *135*, 113–131.

Guida, A., Megreya, A. M., Lavielle-Guida, M., Noël, Y., Mathy, F., van Dijck, J.-P., & Abrahamse, E. (2018). Spatialization in working memory is related to literacy and reading direction: Culture "literarily" directs our thoughts. *Cognition*, *175*, 96–100.

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806.

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences*, *109*(28), 11116–11120.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, *455*(7213), 665–668.

Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology*, *53*(3), 587.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature*, *520*(7546), 220–223.

Hartley, T., Lever, C., Burgess, N., & O'Keefe, J. (2014). Space in the brain: How the hippocampal formation supports spatial cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1635), 20120510.

Hartline, H., & McDonald, P. R. (1947). Light and dark adaptation of single photoreceptor elements in the eye of limulus. *Journal of cellular and comparative physiology*, *30*(3), 225–253.

Harvey, B. M., Klein, B. P., Petridou, N., & Dumoulin, S. O. (2013). Topographic representation of numerosity in the human parietal cortex. *Science*, *341*(6150), 1123–1126.

Hasbroucq, T., & Guiard, Y. (1992). The effects of intensity and irrelevant location of a tactile stimulation in a choice reaction time task. *Neuropsychologia*, *30*(1), 91–94.

Hauser, C. K., & Salinas, E. (2013). Perceptual decision making. In D. Jaeger & R. Jung (Eds.), *Encyclopedia of computational neuroscience* (pp. 1–21). Springer New York.

Hawes, Z., & Ansari, D. (2020). What explains the relationship between spatial and mathematical skills? a review of evidence from brain and behavior. *Psychonomic bulletin & review*, 1–18.

Healey, J. F. (1989). A nabataean sundial from mada'in salih. *Syria. Archéologie, Art et histoire*, *66*(1), 331–336.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Psychology Press.

Hecht, S., Shlaer, S., & Pirenne, M. H. (1942). Energy, quanta, and vision. *Journal of General Physiology*, *25*(6), 819–840.

Hedge, A., & Marsh, N. (1975). The effect of irrelevant spatial correspondences on two-choice response-time. *Acta psychologica*, *39*(6), 427–439.

Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*(7010), 859–862.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of experimental psychology*, *4*(1), 11–26.

Hinrichs, J. V., Yurko, D. S., & Hu, J.-M. (1981). Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(4), 890.

Hommel, B. (1994). Spontaneous decay of response-code activation. *Psychological research*, *56*(4), 261–268.

Hommel, B. (2011). The simon effect as tool and heuristic. *Acta psychologica*, *136*(2), 189–202.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (tec): A framework for perception and action planning. *Behavioral and brain sciences*, *24*(5), 849.

Hubbard, E., Piazza, M., Pinel, P., & Dehaene, S. (2009). Numerical and spatial intuitions: A role for posterior parietal cortex? MIT press.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435–448.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*(1), 106–154.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: Prototype effects in estimating spatial location. *Psychological review*, *98*(3), 352.

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences*, *106*(25), 10382–10385.

Kanayet, F. J., Mattarella-Micke, A., Kohler, P. J., Norcia, A. M., McCandliss, B. D., & McClelland, J. L. (2018). Distinct representations of magnitude and spatial position within parietal cortex during number–space mapping. *Journal of cognitive neuroscience*, *30*(2), 200–218.

Kelly, S. P., & O'Connell, R. G. (2013). Internal and external influences on the rate of sensory evidence accumulation in the human brain. *Journal of Neuroscience*, *33*(50), 19434–19441.

Kiani, R., Corthell, L., & Shadlen, M. N. (2014). Choice certainty is informed by both evidence and decision time. *Neuron*, *84*(6), 1329–1342.

Klam, F., & Graf, W. (2003). Vestibular response kinematics in posterior parietal cortex neurons of macaque monkeys. *European Journal of Neuroscience*, *18*(4), 995–1010.

Kohonen, T. (1972). Correlation matrix memories. *IEEE Trans. Comput., C-21*, 353–359.

Kohonen, T. (1977). *Associative memory: A system theoretical approach.* New York: Springer.

Kornblum, S. (1991). Stimulus-response coding in four classes of stimulus-response ensembles. In J. Requin & G. E. Stelmach (Eds.), *Tutorials in motor neuroscience* (pp. 3–15). Dordrecht, Springer Netherlands.

Kornblum, S. (1994). The way irrelevant dimensions are processed depends on what they overlap with: The case of stroop-and simon-like stimuli. *Psychological research*, *56*(3), 130–135.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility–a model and taxonomy. *Psychological review*, *97*(2), 253.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097–1105.

Kropff, E., Carmichael, J. E., Moser, M.-B., & Moser, E. I. (2015). Speed cells in the medial entorhinal cortex. *Nature*, *523*(7561), 419–424.

Lamberts, K., Tavernier, G., & d'Ydewalle, G. (1992). Effects of multiple reference points in spatial stimulus-response compatibility. *Acta psychologica*, *79*(2), 115–130.

Laming, D. R. J. (1968). Information theory of choice-reaction times.

Lammertyn, J., Notebaert, W., Gevers, W., & Fias, W. (2007). The size of the simon effect depends on the nature of the relevant task. *Experimental psychology*, *54*(3), 202.

Lankford, J. (2013). *History of astronomy: An encyclopedia*. Routledge.

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: An investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438.

Learmount, D., & Norris, G. (1990). Lessons to be learned. *Flig International, 31 October - 6 November*, 24–26.

Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, *72*(1), 246–273.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., & Pitts, W. H. (1959). What the frog's eye tells the frog's brain. *Proceedings of the IRE*, *47*(11), 1940–1951.

Lewald, J., Foltys, H., & Töpper, R. (2002). Role of the posterior parietal cortex in spatial hearing. *Journal of Neuroscience*, *22*(3), RC207–RC207.

Liang, P., & Bose, N. (1996). Neural network fundamentals with graphs, algorithms, and applications. *Mac Graw-Hill*.

Loetscher, T., Bockisch, C. J., Nicholls, M. E., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology*, *20*(6), R264–R265.

Lourenco, S. F., Cheung, C.-N., & Aulet, L. S. (2018). Is visuospatial reasoning related to early mathematical development? a critical review. *Heterogeneity of function in numerical cognition*, 177–210.

Luria, A., Karpov, B., & Yarbuss, A. (1966). Disturbances of active visual perception with lesions of the frontal lobes. *Cortex*, *2*(2), 202–212.

Maiche, A., Budelli, R., & Gomez-Sena, L. (2007). Spatial facilitation is involved in flash-lag effect. *Vision research*, *47*(12), 1655–1661.

Majerus, S., Poncelet, M., Van der Linden, M., Albouy, G., Salmon, E., Sterpenich, V., Vandewalle, G., Collette, F., & Maquet, P. (2006). The left intraparietal sulcus and verbal short-term memory: Focus of attention or serial order? *Neuroimage*, *32*(2), 880–891.

Mante, V., Sussillo, D., Shenoy, K. V., & Newsome, W. T. (2013). Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature*, *503*(7474), 78–84.

Mapelli, D., Rusconi, E., & Umiltà, C. (2003). The snarc effect: An instance of the simon effect? *Cognition*, *88*(3), B1–B10.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information.

Masaki, H., Wild-wall, N., Sangals, J., & Sommer, W. (2004). The functional locus of the lateralized readiness potential. *Psychophysiology*, *41*(2), 220–230.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision making. *Cerebral cortex*, *13*(11), 1257–1269.

McCrink, K., Dehaene, S., & Dehaene-Lambertz, G. (2007). Moving along the number line: Operational momentum in nonsymbolic arithmetic. *Perception & psychophysics*, *69*(8), 1324–1333.

McCrink, K., & Opfer, J. E. (2014). Development of spatial-numerical associations. *Current directions in psychological science*, *23*(6), 439–445.

McCrink, K., & Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychological Science*, *15*(11), 776–781.

McIntosh, R. D., & Schenk, T. (2009). Two visual streams for perception and action: Current trends.

McMillen, T., & Holmes, P. (2006). The dynamics of choice among multiple alternatives. *Journal of Mathematical Psychology*, *50*(1), 30–57.

Mewaldt, S. P., Connelly, C. L., & Simon, J. R. (1980). Response selection in choice reaction time: Test of a buffer model. *Memory & Cognition*, *8*(6), 606–611.

Middlebrooks, P. G., & Sommer, M. A. (2012). Neuronal correlates of metacognition in primate frontal cortex. *Neuron*, *75*(3), 517–530.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge, MA, MIT Press.

Mishkin, M., & Ungerleider, L. G. (1982). Contribution of striate inputs to the visuospatial functions of parieto-preoccipital cortex in monkeys. *Behavioural brain research*, *6*(1), 57–77.

Mizraji, E. (1989). Context-dependent associations in linear distributed memories. *Bulletin of mathematical biology*, *51*(2), 195–205.

Mizraji, E. (2008). Neural memories and search engines. *International Journal of General Systems*, *37*(6), 715–732.

Mizraji, E., & Lin, J. (1997). A dynamical approach to logical decisions. *Complexity*, *2*(3), 56–63.

Mizraji, E., & Lin, J. (2002). The dynamics of logical decisions: A neural network approach. *Physica D: Nonlinear Phenomena*, *168*, 386–396.

Mizraji, E., & Lin, J. (2011). Logic in a dynamic brain. *Bulletin of mathematical biology*, *73*(2), 373–397.

Mizraji, E., Pomi, A., & Valle-Lisboa, J. C. (2009). Dynamic searching in the brain. *Cognitive neurodynamics*, *3*(4), 401–414.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520.

Murata, A., Gallese, V., Luppino, G., Kaseda, M., & Sakata, H. (2000). Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area aip. *Journal of neurophysiology*, *83*(5), 2580–2601.

Najafi, F., & Churchland, A. K. (2018). Perceptual decision-making: A field in the midst of a transformation. *Neuron*, *100*(2), 453–462.

Nasr, K., Viswanathan, P., & Nieder, A. (2019). Number detectors spontaneously emerge in a deep neural network designed for visual object recognition. *Science advances*, *5*(5), eaav7903.

Newcombe, N. S., Möhring, W., & Frick, A. (2018). How big is many? development of spatial and numerical magnitude understanding. In *Heterogeneity of function in numerical cognition* (pp. 157–176). Elsevier.

Newsome, W. T., Britten, K. H., & Movshon, J. A. (1989). Neuronal correlates of a perceptual decision. *Nature*, *341*(6237), 52–54.

Nicoletti, R., & Umiltá, C. (1989). Splitting visual space with attention. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(1), 164.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. *Proceedings of the National Academy of Sciences*, *109*(29), 11860–11865.

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366.

Nieder, A., Freedman, D. J., & Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science*, *297*(5587), 1708–1711.

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences*, *101*(19), 7457–7462.

Niu, M., Impieri, D., Rapan, L., Funck, T., Palomero-Gallagher, N., & Zilles, K. (2020). Receptor-driven, multimodal mapping of cortical areas in the macaque monkey intraparietal sulcus.

Niwa, M., & Ditterich, J. (2008). Perceptual decisions between multiple directions of visual motion. *Journal of Neuroscience*, *28*(17), 4435–4445.

Núñez, R., Cooperrider, K., & Wassmann, J. (2012). Number concepts without number lines in an indigenous group of papua new guinea. *PLoS One*, *7*(4), e35662.

Núñez, R. E. (2011). No innate number line in the human brain. *Journal of Cross-Cultural Psychology*, *42*(4), 651–668.

O'Connell, R. G., Shadlen, M. N., Wong-Lin, K., & Kelly, S. P. (2018). Bridging neural and computational viewpoints on perceptual decision-making. *Trends in neurosciences*, *41*(11), 838–852.

Odic, D., Hock, H., & Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *Journal of Experimental Psychology: General*, *143*(1), 255.

Oh, C., Zak, S. H., & Hui, S. (2005). The generalized brain-state-in-a-box (gbsb) neural network: Model, analysis, and applications.

O'Keefe, J., & Dostrovsky, J. (1971). The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat. *Brain research*.

Opfer, J. E., Kim, D., & Qin, J. (2018). How does the "learning gap" open? a cognitive theory of nation effects on mathematics proficiency. In *Language and culture in mathematical cognition* (pp. 99–130). Elsevier.

Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of vision*, *5*(5), 1–1.

Paré, M., & Dorris, M. C. (2011). The role of posterior parietal cortex in the regulation of saccadic eye movements. *The Oxford handbook of eye movements*, 257–278.

Parker, A. J., & Newsome, W. T. (1998). Sense and the single neuron: Probing the physiology of perception. *Annual review of neuroscience*, *21*(1), 227–277.

Piazza, M., Pinel, P., Le Bihan, D., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305.

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499–503.

Pinhas, M., & Fischer, M. H. (2008). Mental movements without magnitude? a study of spatial biases in symbolic arithmetic. *Cognition*, *109*(3), 408–415.

Pinheiro-Chagas, P., Dotan, D., Piazza, M., & Dehaene, S. (2017). Finger tracking reveals the covert stages of mental arithmetic. *Open Mind*, *1*(1), 30–41.

Pirrone, A., Marshall, J. A., & Stafford, T. (2017). A drift diffusion model account of the semantic congruity effect in a classification paradigm. *Journal of Numerical Cognition*, *3*(1), 77–96.

Pomi, A. (2017). Exploring the sources and mechanisms of cognitive errors in medical diagnosis with associative memory models. *Diagnosis*, *4*(4), 251–259.

Pomi, A., & Mizraji, E. (1999). Memories in context. *BioSystems*, *50*(3), 173–188.

Pomi, A., Mizraji, E., & Lin, J. (2018). Tensor representation of topographically organized semantic spaces. *Neural computation*, *30*(12), 3259–3280.

Pomi, A., & Olivera, F. (2006). Context-sensitive autoassociative memories as expert systems in medical diagnosis. *BMC medical informatics and decision making*, *6*(1), 1–11.

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3–25.

Pratte, M. S., Rouder, J. N., Morey, R. D., & Feng, C. (2010). Exploring the differences in distributional properties between stroop and simon effects using delta plots. *Attention, Perception, & Psychophysics*, *72*(7), 2013–2025.

Priftis, K., Zorzi, M., Meneghello, F., Marenzi, R., & Umiltà, C. (2006). Explicit versus implicit processing of representational space in neglect: Dissociations in accessing the mental number line. *Journal of cognitive neuroscience*, *18*(4), 680–688.

Proctor, R. W., & Lu, C.-H. (1994). Referential coding and attention-shifting accounts of the simon effect. *Psychological research*, *56*(3), 185–195.

Proctor, R. W., Lu, C.-H., & Van Zandt, T. (1992). Enhancement of the simon effect by response precuing. *Acta Psychologica*, *81*(1), 53–74.

Proctor, R. W., & Schneider, D. W. (2018). Hick's law for choice reaction time: A review. *Quarterly Journal of Experimental Psychology*, *71*(6), 1281–1299.

Proctor, R. W., & Vu, K.-P. L. (2006). *Stimulus-response compatibility principles: Data, theory, and application*. CRC press.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., & Schall, J. D. (2012). Response variability of frontal eye field neurons modulates with sensory input and saccade preparation but not visual search salience. *Journal of neurophysiology*, *108*(10), 2737–2750.

Putnam, H. (1975). Philosophy and our mental life. In *Philosophical papers* (pp. 291–303). Cambridge University Press.

Ranzini, M., Lisi, M., & Zorzi, M. (2016). Voluntary eye movements direct attention on the mental number space. *Psychological research*, *80*(3), 389–398.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological review*, *111*(2), 333.

Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive development*, *13*(4), 453–478.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2p1), 274.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, *25*(1), 31–40.

Robson, E. (2002). Words and pictures: New light on plimpton 322. *The American Mathematical Monthly*, *109*(2), 105–120.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionst model of decision making. *Psychological review*, *108*(2), 370.

Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *Journal of neuroscience*, *22*(21), 9475–9489.

Ronayne, D., & Brown, G. D. (2017). Multi-attribute decision by sampling: An account of the attraction, compromise and similarity effects. *Journal of Mathematical Psychology*, *81*, 11–27.

Rorie, A. E., Gao, J., McClelland, J. L., & Newsome, W. T. (2010). Integration of sensory and reward information during perceptual decision-making in lateral intraparietal cortex (lip) of the macaque monkey. *PloS one, 5*(2), e9308.

Rosenbloom, P. S., & Newell, A. (1988). An integrated computational model of stimulus-response compatibility and practice. In *Psychology of learning and motivation* (pp. 1–52). Elsevier.

Rosenbloom, P. (1986). The chunking of goal hierarchies: A model of stimulus–response compatibility and practice. *Universal subgoaling and chunking: The automatic generation and learning of goal hierarchies*, 133–282.

Rottermann, A., & Vu, K.-P. L. (2009). Reversing the simon effect with prior practice of noncorresponding location words, In *Symposium on human interface*. Springer.

Rugani, R., Fontanari, L., Simoni, E., Regolin, L., & Vallortigara, G. (2009). Arithmetic in newborn chicks. *Proceedings of the Royal Society B: Biological Sciences, 276*(1666), 2451–2460.

Rugani, R., Kelly, D. M., Szelest, I., Regolin, L., & Vallortigara, G. (2010). Is it only humans that count from left to right? *Biology Letters, 6*(3), 290–292.

Rugani, R., Regolin, L., & Vallortigara, G. (2007). Rudimental numerical competence in 5-day-old domestic chicks (gallus gallus): Identification of ordinal position. *Journal of Experimental Psychology: Animal Behavior Processes, 33*(1), 21.

Rugani, R., Vallortigara, G., Priftis, K., & Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Science, 347*(6221), 534–536.

Rugani, R., Vallortigara, G., & Regolin, L. (2013). Numerical abstraction in young domestic chicks (gallus gallus). *PLoS One, 8*(6), e65262.

Sagiv, N., Simner, J., Collins, J., Butterworth, B., & Ward, J. (2006). What is the relationship between synaesthesia and visuo-spatial number forms? *Cognition, 101*(1), 114–128.

Salzer, Y., de Hollander, G., & Forstmann, B. U. (2017). Sensory neural pathways revisited to unravel the temporal dynamics of the simon effect: A model-based cognitive neuroscience approach. *Neuroscience & Biobehavioral Reviews, 77*, 48–57.

Sarton, G. (1937). A hindu decimal ruler of the third millennium (second note). *Isis, 26*(2), 304–305.

Sasanguie, D., De Smedt, B., & Reynvoet, B. (2017). Evidence for distinct magnitude systems for symbolic and non-symbolic number. *Psychological research, 81*(1), 231–242.

Schall, J. D. (2002). The neural selection and control of saccades by the frontal eye field. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 357*(1424), 1073–1082.

Schmiedek, F., Oberauer, K., Wilhelm, O., Suess, H.-M., & Wittmann, W. W. (2007). Individual differences in components their relations to working of reaction time distributions and memory and intelligence. *Journal of Experimental Psychology-General, 136*(3), 414–429.

Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development*, 630–633.

Sella, F., Lucangeli, D., & Zorzi, M. (2020). The interplay between spatial ordinal knowledge, linearity of number-space mapping, and arithmetic skills. *Cognitive Development, 55*, 100915.

Servant, M., White, C., Montagnini, A., & Burle, B. (2016). Linking theoretical decision-making mechanisms in the simon task with electrophysiological data: A model-based neuroscience study in humans. *Journal of Cognitive Neuroscience, 28*(10), 1501–1521.

Servant, M., White, C., Montagnini, A., & Burle, B. (2015). Using covert response activation to test latent assumptions of formal decision-making models in humans. *Journal of Neuroscience*, *35*(28), 10371–10385.

Shadlen, M. N., & Newsome, W. T. (2001). Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of neurophysiology*, *86*(4), 1916–1936.

Shaki, S., & Fischer, M. H. (2008). Reading space into numbers–a cross-linguistic comparison of the snarc effect. *Cognition*, *108*(2), 590–599.

Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the snarc effect. *Psychonomic bulletin & review*, *16*(2), 328–331.

Shepard, R. N., Kilpatric, D. W., & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive psychology*, *7*(1), 82–138.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological science*, *14*(3), 237–250.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of experimental psychology*, *81*(1), 174.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In *Advances in psychology* (pp. 31–86). Elsevier.

Simon, J. R., & Craft, J. L. (1970). Effects of an irrelevant auditory stimulus on visual choice reaction time. *Journal of experimental psychology*, *86*(2), 272.

Simon, J. R., & Rudell, A. P. (1967). Auditory sr compatibility: The effect of an irrelevant cue on information processing. *Journal of applied psychology*, *51*(3), 300.

Simon, J. R., & Small, A. J. (1969). Processing auditory information: Interference from an irrelevant cue. *Journal of Applied Psychology*, *53*(5), 433.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial intelligence*, *46*(1-2), 159–216.

Sotiropoulos, G., Seitz, A. R., & Seriès, P. (2014). Contrast dependency and prior expectations in human speed perception. *Vision research*, *97*, 16–23.

Spaepen, E., Coppola, M., Spelke, E. S., Carey, S. E., & Goldin-Meadow, S. (2011). Number without a language model. *Proceedings of the National Academy of Sciences*, *108*(8), 3163–3168.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.

Stocco, A., Murray, N. L., Yamasaki, B. L., Renno, T. J., Nguyen, J., & Prat, C. S. (2017). Individual differences in the simon effect are underpinned by differences in the competitive dynamics in the basal ganglia: An experimental verification and a computational model. *Cognition*, *164*, 31–45.

Stone, M. (1960). Models for choice-reaction time. *Psychometrika*, *25*(3), 251–260.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of experimental psychology*, *18*(6), 643.

Summerfield, C., & Blangero, A. (2017). Perceptual decision-making: What do we know, and what do we not know? In *Decision neuroscience* (pp. 149–162). Elsevier.

Summerfield, C., & Tsetsos, K. (2012). Building bridges between perceptual and economic decision-making: Neural and computational mechanisms. *Frontiers in neuroscience*, *6*, 70.

Tagliabue, M., Zorzi, M., Umiltà, C., & Bassignani, F. (2000). The role of long-term-memory and short-term-memory links in the simon effect. *Journal of Experimental Psychology: Human perception and performance*, *26*(2), 648.

Taube, J. S., Muller, R. U., & Ranck, J. B. (1990). Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis. *Journal of Neuroscience*, *10*(2), 420–435.

Tolman, E. C., Ritchie, B., & Kalish, D. (1946). Studies in spatial learning. i. orientation and the short-cut. *Journal of experimental psychology*, *36*(1), 13.

Trueblood, J. S., Brown, S. D., & Heathcote, A. (2014). The multiattribute linear ballistic accumulator model of context effects in multialternative choice. *Psychological review*, *121*(2), 179.

Tsetsos, K., Gao, J., McClelland, J. L., & Usher, M. (2012). Using time-varying evidence to test models of decision dynamics: Bounded diffusion vs. the leaky competing accumulator model. *Frontiers in neuroscience*, *6*, 79.

Tsetsos, K., Usher, M., & Chater, N. (2010). Preference reversal in multiattribute choice. *Psychological review*, *117*(4), 1275.

Tversky, A. (1972). Elimination by aspects: A theory of choice. *Psychological review*, *79*(4), 281.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458.

Ujfalussy, D. J., Miklósi, Á., & Bugnyar, T. (2013). Ontogeny of object permanence in a non-storing corvid species, the jackdaw (corvus monedula). *Animal cognition*, *16*(3), 405–416.

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive psychology*, *78*, 148–174.

Umiltà, C., Priftis, K., & Zorzi, M. (2009). The spatial representation of numbers: Evidence from neglect and pseudoneglect. *Experimental Brain Research*, *192*(3), 561–569.

Urcuioli, P. J., Vu, K.-P. L., & Proctor, R. W. (2005). A simon effect in pigeons. *Journal of Experimental Psychology: General*, *134*(1), 93.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological review*, *108*(3), 550.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological review*, *111*(3), 757.

Usher, M., Olami, Z., & McClelland, J. L. (2002). Hick's law in a stochastic race model with speed–accuracy tradeoff. *Journal of Mathematical Psychology*, *46*(6), 704–715.

Valle-Inclán, F. (1996). The locus of interference in the simon effect: An erp study. *Biological Psychology*, *43*(2), 147–162.

Valle-Lisboa, J. C., Pomi, A., Cabana, A., Elvevåg, B., & Mizraji, E. (2014). A modular approach to language production: Models and facts. *Cortex*, *55*, 61–76.

Vallortigara, G. (2017). 2.2 an animal's sense of number. *The nature and development of mathematics: Cross disciplinary perspectives on cognition, learning and culture*, 43–65.

Van Opstal, F., Gevers, W., De Moor, W., & Verguts, T. (2008). Dissecting the symbolic distance effect: Comparison and priming effects in numerical and nonnumerical orders. *Psychonomic Bulletin & Review*, *15*(2), 419–425.

van Dijck, J.-P., Abrahamse, E. L., Majerus, S., & Fias, W. (2013). Spatial attention interacts with serial-order retrieval from verbal working memory. *Psychological Science*, *24*(9), 1854–1859.

van Dijck, J.-P., & Fias, W. (2011). A working memory account for spatial–numerical associations. *Cognition*, *119*(1), 114–119.

Verguts, T., & Fias, W. (2004). Representation of number in animals and humans: A neural model. *Journal of cognitive neuroscience*, *16*(9), 1493–1504.

Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation. *Psychonomic bulletin & review*, *12*(1), 66–80.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*(1), 37–58.

Vickers, D., & Packer, J. (1982). Effects of alternating set for speed or accuracy on response time, accuracy and confidence in a unidimensional discrimination task. *Acta psychologica*, *50*(2), 179–197.

Wallace, R. J. (1971). Sr compatibility and the idea of a response code. *Journal of experimental psychology*, *88*(3), 354.

Wallis, J. (1685). *A treatise of algebra*. Davis.

Wang, H., & Proctor, R. W. (1996). Stimulus–response compatibility as a function of stimulus code and response modality. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1201.

Wang, X.-J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*(5), 955–968.

Ward, C., & Smuts, B. B. (2007). Quantity-based judgments in the domestic dog (canis lupus familiaris). *Animal cognition*, *10*(1), 71–80.

Wee, C. (2016). 5 ibn al-haytham and the experimental method. In *The bright dark ages* (pp. 93–108). Brill.

White, C., Servant, M., & Logan, G. (2017). Practical considerations for using conflict-based diffusion models to interpret choice rt data. *Psychonomic Bulletin & Review*.

Whitelaw, I. (2007). *A measure of all things: The story of man and measurement*. Macmillan.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta psychologica*, *41*(1), 67–85.

Wiegand, K., & Wascher, E. (2005). Dynamic aspects of stimulus-response correspondence: Evidence for two mechanisms involved in the simon effect. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(3), 453.

Wilkey, E. D., & Ansari, D. (2019). Challenging the neurobiological link between number sense and symbolic numerical abilities. *Ann. NY Acad. Sci*, *40*, 1–23.

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience & Biobehavioral Reviews*, *57*, 209–219.

Woods, D. M., Ketterlin Geller, L., & Basaraba, D. (2018). Number sense on the number line. *Intervention in School and Clinic*, *53*(4), 229–236.

Wyart, V., De Gardelle, V., Scholl, J., & Summerfield, C. (2012). Rhythmic fluctuations in evidence accumulation during decision making in the human brain. *Neuron*, *76*(4), 847–858.

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11.

Zhang, H. H., Zhang, J., & Kornblum, S. (1999). A parallel distributed processing model of stimulus–stimulus and stimulus–response compatibility. *Cognitive psychology*, *38*(3), 386–432.

Zorzi, M., & Butterworth, B. (1999). A computational model of number comparison. LEA.

Zorzi, M., Priftis, K., Meneghello, F., Marenzi, R., & Umiltà, C. (2006). The spatial representation of numerical and non-numerical sequences: Evidence from neglect. *Neuropsychologia*, *44*(7), 1061–1067.

Zorzi, M., Priftis, K., & Umiltà, C. (2002). Neglect disrupts the mental number line. *Nature*, *417*(6885), 138–139.

Zorzi, M., Stoianov, I., & Umiltà, C. (2005). Computational modeling of numerical cognition. *Handbook of mathematical cognition*, *19*, 67–84.

Zorzi, M., & Umiltá, C. (1995). A computational model of the simon effect. *Psychological Research, 58*(3), 193–205.