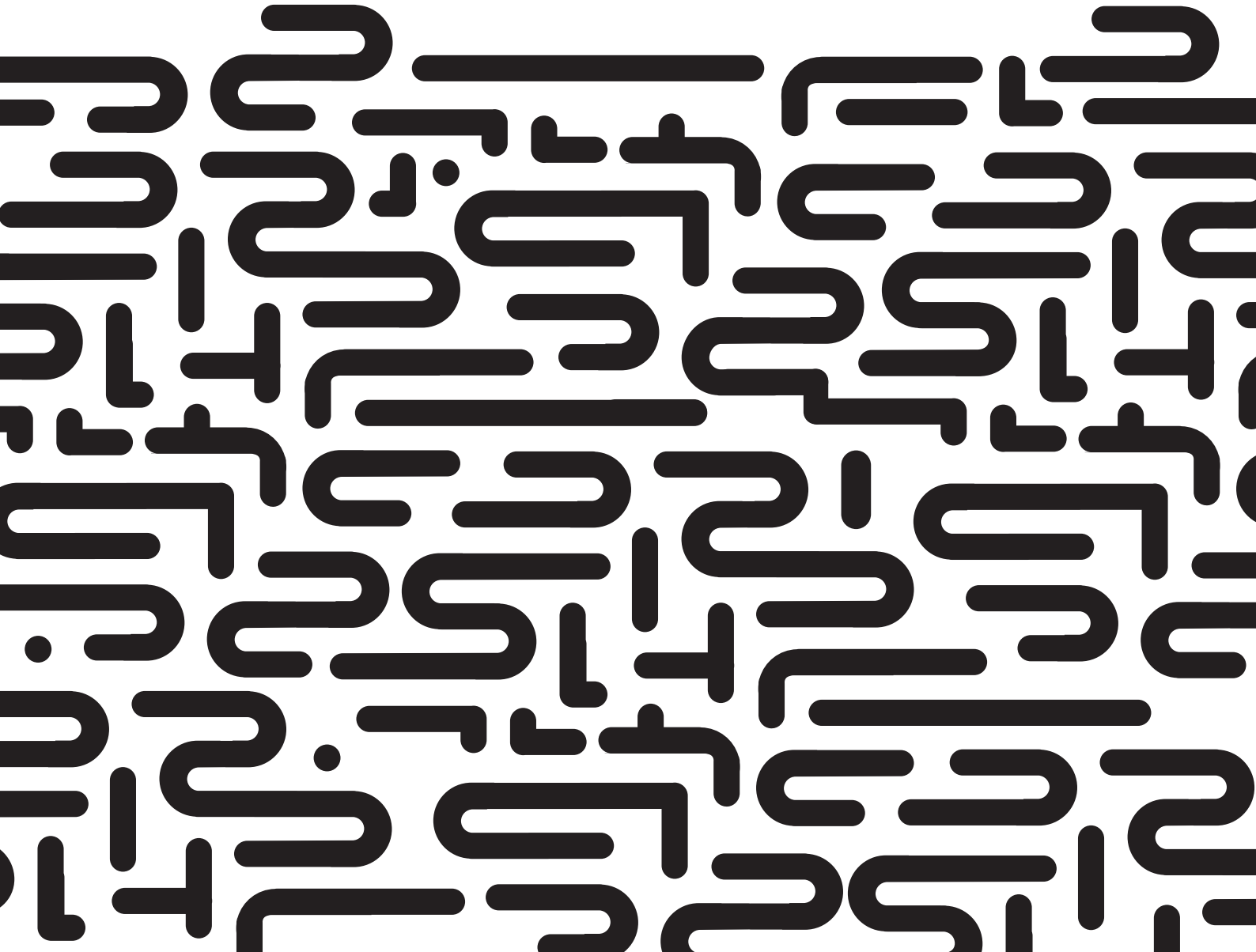
# The Representation Of Numerosity In The Human Brain And Machines

Alireza Karami



UNIVERSITY OF TRENTO

DOCTORAL THESIS

# The representation of numerosity in the human brain and machines

*Author:*
Alireza KARAMI

*Supervisor:*
Prof. Manuela PIAZZA

*A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy*

*in the*

Perception to Concepts (per2con)
Center for Mind/Brain Sciences (CIMeC)

February 15, 2024

# Declaration of Authorship

I, Alireza KARAMI, declare that this thesis titled, "The representation of numerosity in the human brain and machines" and the content therein are my own work. The following list enumerates the publications that will be extracted from my thesis or from side projects conducted during my doctoral program.

- Neural codes for visual numerosity independent of other quantities are present both in the dorsal and in the ventral stream of the human brain; *Alireza Karami, Elisa Castaldi, Evelyn Eger, Manuela Piazza* Under revision.
- Fast, direct, and parallel extraction of visual numerosity in the ventral and dorsal stream of the human brain; *Alireza Karami, Elisa Castaldi, Evelyn Eger, Martin Hebart, Manuela Piazza* In preparation.
- Numerosity representations in CCNs do not well approximate those observed in the human brain; *Alireza Karami, Elisa Castaldi, Evelyn Eger, Shahriyar Noei, Manuela Piazza* In preparation.
- Investigating representation of multiple objects in human brain through naturalistic stimuli; *Alireza Karami, Marius Peelen, Manuela Piazza* In preparation.
- The Brain Basis of Mathematical Difficulties: A Review of Neuroimaging Literature; *Paula A. Maldonado Moscoso, Alireza Karami, Manuela Piazza* Under revision.

Signed:

Date:

UNIVERSITY OF TRENTO

# *Abstract*

Department of Psychology and Cognitive Science

Center for Mind/Brain Sciences (CIMeC)

Doctor of Philosophy

**The representation of numerosity in the human brain and machines**

by Alireza KARAMI

The capacity to estimate the number of objects (numerosity) in the environment is ontogenetically precocious and phylogenetically ancient. In animals, this ability holds significant adaptive advantages, directly influencing survival and reproductive success. In humans, it may serve an additional purpose by providing a start-up kit for the acquisition of symbolic numbers, thus making it a potential focus for mathematics education and intervention strategies. Behavioral, neurophysiological, and neuroimaging findings suggest that numerosity information is directly extracted from the environment. However, numerosity is inherently linked with other visual characteristics of sets (such as larger sets often occupy more space or are more densely spaced), making it challenging to determine the extent to which the observed response to numerosity is distinct from the response to other visual attributes.

In my PhD research I provide experimental evidence through neuroimaging and computational modeling techniques elucidating where, when, and how numerical information is encoded in the human brain. This work therefore provides a threefold contribution. First, I show that numerosity is represented over and above non-numeric visual features in a widespread network of areas starting from early visual areas and further amplified in associative areas along the dorsal but also notably the ventral stream, and that the neural representational geometries of regions across the two steams are substantially identical. Second, I showed that numerosity is represented at an early stage and seemingly in parallel across of a set of regions including early visual, parietal, and temporal, preceding the emergence of non-numeric features that could indirectly contribute to numerosity computation. Finally, by comparing the fMRI data with a convolutional neural network (CNN) to explore similarities and differences between the model and human brain data, I discovered that although the CNN can perform approximate numerosity comparisons and the

structure of their representation in their hidden layers captures well numerosity representation in early visual areas of humans, it falls short of fully simulating the way in which associative brain regions represent numerosity.

Taken together, the findings of this thesis provide experimental evidence supporting the notion that number is a primary visual feature, encoded independent from other visual features quickly and widely across the human brain. Furthermore, they emphasize the need for additional investigation to unravel the computational mechanisms underlying numerosity in the human brain.

# *Acknowledgements*

I would like to express my gratitude to all those who have contributed to the completion of my Ph.D. thesis.

First and foremost, I am deeply thankful to my advisor, Manuela Piazza, for her mentorship throughout my four-year Ph.D. journey. From her, I learned that in science, mastering the art of simplifying complex ideas is essential.

I also express gratitude to my colleagues and researchers, including Martin Hebart, Evelyn Eger, Elisa Castaldi, Scott Laurence Fairhall, Stefania Bracci, Marco Buiatti, Paula Andrea Maldonado Moscoso, Marie Amalric, Shahryar Noei, Simone Viganò, Marco Zorzi, Alberto Testolin, Mahdi Khaligh-Razavi, and Marius Peelen from whom I have gained valuable insights through the exchange of ideas and collaborative efforts. I would like to express my gratitude to Donatella Zingaro and Anna Burato for their assistance in a portion of the data collection. Additionally, I extend my thanks to the CIMeC neuroimaging labs technicians, Gianpiero Monittola, Manuela Orsini, Nicola Pace, Stefano Tambalo, Ilaria Mirandola, and Davide Tabarelli, who provided support throughout the process of neuroimaging data collection.

My heartfelt thanks go to my family, particularly my father and my mother, for their unconditional support and encouragement. I would also like to express my gratitude to my wife, Sedighe, who joined me in the midst of my Ph.D. journey, offering invaluable encouragement and love. Completing this journey without her alone, if not impossible, would have been exceptionally challenging.

Last but not least, I must express my gratitude to Mohammad Ardeshir, a logician whom I encountered during my bachelor's studies in my home country, Iran. Attending his class on logic and the foundations of mathematics sparked my curiosity about mathematics. Even though I eventually chose a different path to study mathematics, I believe that without his influence, I might never have ventured to Italy to explore how numbers are represented in the human mind. Perhaps, I would have pursued a career as a computer programmer in a company.

# *Open Data Statement*

Conducting an experimental study undoubtedly entails numerous practical steps. It's not an exaggeration to state that completing this four-year thesis would have been impossible without the availability of open tools. I am deeply grateful to all the individuals dedicated to providing their code for the research community's benefit, contributing significantly to the advancement of science. In the spirit of collaboration, I am committed to sharing as much of my own code, experiments, analyses, and data as possible.

For the design of my thesis, my wife drew inspiration from the brickwork found in Iranian architecture, seamlessly integrating it with the curved structure of the human brain cortex.

Khosrogerd Minaret in *Sabzevar, Iran*

# Contents

# List of Figures

# List of Abbreviations

| | |
|---|---|
| **ANS** | **A**pproximate **N**umber **S**ystem |
| **OTS** | **O**bject **T**racking **S**ystem |
| **IPS** | **I**ntra**p**arietal **S**ulcus |
| **ITG** | **Inferior T**emporal **G**yrus |
| **NFA** | **N**umber **F**orm **A**rea |
| **VNFA** | **V**isual **N**umber **F**orm **A**rea |
| **RSA** | **R**epresentation **S**imilarity **A**nalysis |
| **RDM** | **R**epresentational **D**isimilarity **M**atrix |
| **MDS** | **M**ulti**d**imensional **S**caling |
| **CNN** | **C**onvolutional **N**eural **N**etwork |
| **GCSS** | **G**roup-**C**onstrained **S**ubject-**S**pecific |

# CHAPTER 1

## General Introduction

# Chapter 1

# General Introduction

**SOCRATES** *Now tell me, my young friend, what is the object of mathematics? What things does a mathematician study?*

**HIPPOCRATES** *I have asked Theaitetos the same question. He answered that a mathematician studies numbers and geometrical forms.*

**SOCRATES** *Well, the answer is right, but would you say that these things exist?*

**HIPPOCRATES** *Of course. How can we speak of them if they do not exist?*

**SOCRATES** *Then tell me, if there were no mathematicians, would there be prime numbers, and if so, where would they be?*

---

*A Socratic Dialogue on Mathematics*

*ALFRÉD RÉNY*

I was born in Tehran (Figure 1.1), the capital of Iran. As the capital of Iran, Tehran is a big and highly populated city with a population density of 11,800/km$^2$. I traveled over 4,669.6 km from Iran to Italy for my Ph.D. Currently, I live in Trento (Figure 1.1). Trento is a city in the north of Italy with beautiful nature and eye-catching scenery. Compared to Tehran, Trento, with a population density of 760/km$^2$, is small. Tehran and Trento may not have much in common, as one may expect by looking at their photos. However, all four images, as may one notice immediately, have something in common. All of them contain multiple objects: people, cars, trees, or buildings.

Numbers are prevalent in everyday life, shaping a wide array of activities such as estimating the number of people or cars in the street or communicating the population of a city or the distance between two countries. Research indicates that difficulties in numeracy negatively affect mental well-being and employment rates (Parsons & Bynner, 2005). Moreover, the inability to grasp essential numeracy skills during primary education can impose expenses on the public purse. However, timely and effective numeracy interventions can reduce these costs (J. Gross et al., 2008). Currently we miss a complete model of numerical cognition that could effectively guide both education and remediation.


FIGURE 1.1: Images from (A) Tehran: Azadi Stadium and Tabiat Bridge, and (B) Trento: Cathedral Square and the Italian Garden of Buonconsiglio.

Perceiving and manipulating numbers is not only vital for our daily lives but has also captivated the minds of philosophers for centuries. Philosophers of mathematics are preoccupied with two fundamental questions: the nature of numbers (ontology) and how we can access them (epistemology) (Horsten, 2023). Philosophers are divided into two main camps: anti-realists who argue that numbers do not exist, while realists assert their existence. Within the realist, there are at least three different views: Numbers, according to some realists, are solely mental entities, while others contend that numbers are aspects of the physical universe that exist outside of people's brains (Balaguer, 1998).

For a thorough discussion on the matter, it's crucial to clarify that the term "number" in the preceding paragraphs actually refers to two distinct concepts. Firstly, it can indicate the property of any set of objects that specifies its cardinal value, often termed "numerosity" or "howmanyness." Brains are capable of perceiving and representing this property. In this sense, non-symbolic number representations fall under the realm of percepts—for example, number of people or cars in the street or the count of taps against a table. Secondly, the term "number" can also denote the symbols devised by humans to quantify, measure, and order. For instance, we can use numbers for measuring distance or population.

In the cognitive science literature, the ability to represent non-symbolic numbers

(numerosities) is commonly referred to as "number sense." Tobias Danzig was the first to coin this term, defining it as an individual's ability to detect changes in a small collection when objects are added or removed without the person's direct awareness (Dantzig & Mazur, 1967). Over time, the term "number sense" has broadened and it currently mainly denotes the capacity to estimate the number of objects in sets, irrespective of the set size. There is evidence that number sense is ontogenetically precocious, as infants, shortly after birth spontaneously differentiate sets on the bases of their numerosity (Hyde et al., 2010; V. R. Izard et al., 2009; V. Izard et al., 2008; Xu & Spelke, 2000). Moreover, the ability to perceive the number of objects in the natural environment carries evolutionary advantages for non-human animals (Nieder, 2020). This phylogenetically ancient ability has been observed in birds (Ditz & Nieder, 2016; Lyon, 2003; Rugani et al., 2008, 2009, 2015; Scarf et al., 2011; Templeton et al., 2005; Vallortigara, 2017), fish (Agrillo & Bisazza, 2018; Agrillo et al., 2012; Dadda et al., 2009; Hager & Helfman, 1991; Potrich et al., 2015, 2019), insects (Carazo et al., 2009; Dacke & Srinivasan, 2008; Giurfa, 2019; H. Gross et al., 2009; MaBouDi et al., 2020; Skorupski et al., 2018), amphibians (Balestrieri et al., 2019; Stancher et al., 2014; Uller et al., 2003), rodents (Davis & Albert, 1986), cats (Pisa & Agrillo, 2008), elephants (Perdue et al., 2012), and primates (Beran, 2004, 2012; Beran et al., 2011; Jordan et al., 2005, 2008; Thomas & Chase, 1980). Additionally, members of oral indigenous communities with no formal education systems and no verbal counting routines, such as the Pirahã (Gordon, 2004), and the Munduruku tribes (Pica et al., 2004), both residing in the Amazon rainforest of Brazil , exhibit the ability to represent and manipulate numerosity approximately.

Depending on the numerosity of the set and the attentional resources available, two systems can be at play in enumeration: the Object Tracking System (OTS), also known as the parallel individuation system, and the Approximate Number System (ANS). The OTS only handles small sets of one to about four items, whereas the ANS enables us to grasp the cardinality of an unlimited number of objects. However, while the OTS represents numbers precisely, the ANS representation is noisy and, as the name suggests, approximate. The ANS behavioral hallmarks are the "numerical distance effect" and the "numerical magnitude effect". According to the numerical distance effect, it is easier to discriminate between distant numbers than those that are closer. For instance, discriminating between 7 and 10 dots is easier than 7 and 8. On the other hand, based on the numerical magnitude effect for equal numerical distances as the numbers to be compared increase in magnitude, performance declines. For instance, it is easier to differentiate between 4 and 5 dots than between 12 and 13. To achieve a discrimination performance comparable to that seen in the case of 4 versus 5, the numerical distance would need to increase proportionally with the magnitudes of the numbers, in this example to 12 versus 15. Therefore, the capacity to discriminate between quantities is influenced by their ratio and is thus "ratio-dependent" which is the reason why numerosity is accounted for by Weber's law(Dehaene, 2011; Nieder, 2019).

To represent the quantity of objects, the OTS individuates objects by placing them in mental files. As each mental file can simultaneously and rapidly hold one object, the reaction time for discriminating between one to four items remains relatively constant (Nieder, 2019). This phenomenon, often referred to as subitizing (from the Latin word "subitus," meaning "immediately"), was termed by Kaufman and Lord (1949) and refers to OTS capacity.

Current cognitive development theories propose that knowledge acquisition during development relies on a few fundamental "core knowledge" systems (Spelke & Kinzler, 2006). Number sense is thought to be one of these core knowledge systems, playing a crucial role in supporting culture-based learning of mathematics, particularly arithmetic. Numerous studies have demonstrated that individuals with a more acute [1] number sense exhibits more skilled calculation abilities (for a meta-analysis see: Chen & Li, 2014; Schneider et al., 2016), and that number sense acuity is lower in children with developmental dyscalculia compared to age and intelligence-matched controls (Piazza et al., 2010). Compared with the OTS, the evidence, albeit sometimes contradictory (for a review see: De Smedt et al., 2013) suggests that the ANS plays a more fundamental role in the development of mathematical abilities (for a review see: Piazza, 2010).

The most common task to measure ANS acuity is the dot comparison task (Figure 1.2). In this task, participants are required to choose the more numerous (or, left often, the less numerous) sets of pairs of dots following a brief presentation, either simultaneous or sequential. Similar to other physical magnitudes (such as duration or weight), non-symbolic numbers are represented with an internal variability that scales with stimulus magnitude (here cardinality), a phenomenon described by Weber's law (Whalen et al., 1999). Although traditional models for estimating numerical acuity, often expressed as the Weber fraction (Barth et al., 2006; Dehaene, 2003; Piazza et al., 2004; Pica et al., 2004), approach the question of the relation between numerosity estimation and the other visual features of the arrays (e.g. total surface, density) by randomizing the visual features across trials and numerosities when generating the dot stimuli for the experiments, recent evidence indicates that the visual features may interfere in numerical judgements in a more profound way than initially thought, and thus suggest that it is better to model, rather than try to control, those factors in the analyses (DeWind et al., 2015; Gebuis & Reynvoet, 2012; Piazza et al., 2018). The underlying source of such interference is currently unknown, but researchers have proposed at least four potential explanations. First, it's conceivable that both number, as well as co-varying visual parameters, like dot size or density are encoded through extremely low-level visual features such as contrast or spatial

[1] When referring to acuity, I mean how imprecise the internal vision of numbers is. For instance, individuals with a more acute number sense might estimate a set of 15 items as having 13, 14, 16, or 17 items. In contrast, individuals with a less acute number sense might estimate the same set as containing between 11 to 22 items. Similarly, subjects with a more acute number sense may be able to differentiate sets differing by a ratio of 9:10, while those with a less acute number sense may be able to differentiate sets only if they differ by a 5:6 ratio.

frequency (Dakin et al., 2011; Stoianov & Zorzi, 2017) but see (Burr & Ross, 2008) for a different view. According to this perspective, altering size can lead to a change in perceived numerosity, but the underlying reason for this phenomenon is that both size and numerosity are derived from a common low-level feature. The second possibility is that numerical information is computed through the combination of other quantities: mathematically number can be computed by multiplying total field area of dots by their density as well as dividing the total surface area of dots by their average item area. According to this view, because the different visual features concur in the computation of number, their estimation directly impacts the estimation of numerosity. A third hypothesis proposes that numeric and non-numeric features are independently extracted by separate direct channels but compete for the same decision-making component, akin to the classical STROOP effect. This hypothesis suggests that interference arises at the stage of response selection (Barth, 2008; Hurewitz et al., 2006; Nys & Content, 2012; Odic et al., 2016; Piazza et al., 2018; Rousselle & Noël, 2008; Van Opstal & Verguts, 2013). A final view, synthetized in the ATOM (as in "A Theory Of Magnitude") theory (Walsh, 2003) suggests that numerical information is represented on a shared representational scale along other visual features, as suggested by partially overlapping representations of different quantities in the parietal cortex (Harvey et al., 2015; Tudusciuc & Nieder, 2007, 2009). The nature of this representation is unitless, and various types of magnitudes are represented on a domain-general magnitude scale. This scale encodes only the distinction between more and less (Cantrell & Smith, 2013; Lourenco & Longo, 2010; Sheahan et al., 2021).

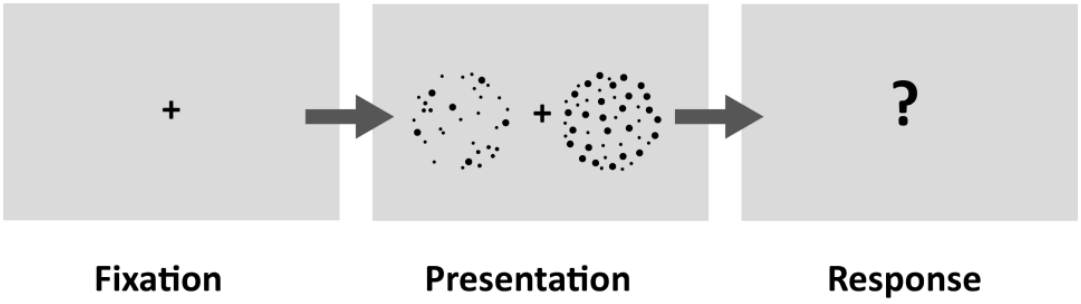



FIGURE 1.2: Numerosity comparison task

The mechanism by which the brain extracts the cardinality of a set is a topic of ongoing debate. It has been suggested that evolution shapes human numerical concepts (for reviews see: Cantlon, 2017). According to this hypothesis, primate perception is object-oriented. In the process of making quantitative judgments, the perceptual systems of primates might be predisposed to prioritize the numerosity associated with discrete objects, as opposed to dimensions such as area and size that are based on surface characteristics (Ferrigno et al., 2017). This is evidenced by a greater allocation of primate cortex to objects compared to surfaces (Cantrell & Smith, 2013).

Despite a growing body of evidence from behavioral, neurophysiological, and neuroimaging studies suggesting that numerosity is estimated through a dedicated neural pathway (for reviews see: Eger, 2016; Nieder, 2016; Piazza & Eger, 2016), various combinations of co-varying non-numerical factors, such as item size or density, also appear to significantly influence estimation in both adults and children (for a review see: Leibovich et al., 2016).

In what follows, I will briefly review some of the most important studies within behavioral, neurophysiological, and neuroimaging literature that have been conducted to investigate the neural representation of numerosity and show that most support the idea of a direct representation of numerosity in the brain. Additionally, I will also address the objections that were raised against these studies. Lastly, I will review notable computational models proposed for extracting numerosity information from a visual set of dots. I will assess whether, according to these models, numerosity information is extracted directly through a dedicated mechanism or indirectly through other non-numeric features.

## 1.1 Behavior

The numerical abilities of babies have been studied extensively using visual, auditory, and cross-modal stimulation. Collectively, these studies have led to a consensus that infants possess an amodal or supramodal approximate number system for representing discrete quantities measurable soon after birth. Initial investigations into infants' numerical abilities employed habituation procedures (Antell & Keating, 1983; Starkey & Cooper, 1980; Strauss & Curtis, 1981; Xu & Spelke, 2000). In these studies, infants are habituated to images of various numerosities (e.g., 2 or 3 dots) and then presented with novel quantities. If habituated to two dots, they exhibit dishabituation to three, and vice versa, indicating their ability to detect changes in cardinality. This ability to dishabituate to novel stimuli has also been observed with auditory stimuli (Lipton & Spelke, 2003). In both visual and auditory presentations of numerosity, there are non-numeric confounds that cannot be simultaneously controlled. Infants' numerical abilities have also been explored using a preferential looking paradigm, where they gaze longer at the visual quantity matching the auditory quantity (Coubart et al., 2013; V. R. Izard et al., 2009), for a review see Cantrell and Smith (2013). This cross-modal paradigm is truly devoid of sensory factors confounded with numerosity, thus it firmly demonstrates the existence of a mode-invariant system for extracting purely numerical information.

Behavioral evidence from adaptation studies suggests that number is perceived directly through a dedicated mechanism, thus that it can be considered (similarly to colour) a "primary" visual property of the image. Adaptation causes the change in the appearance of a stimulus following exposure to that stimulus for several seconds. The susceptibility of a perceptual mechanism to adaptation is indeed typically

taken as evidence of the presence of a dedicated perceptual system (P. Thompson & Burr, 2009). Burr and Ross (2008) showed that visual numerosity is prone to adaptation: the number of dots in a very numerous visual stimulus was underestimated after several seconds of exposure; similarly, the number of dots in a not very numerous visual stimulus was overestimated. Despite this striking effect, whether numerosity can be considered a primary visual feature was contested by Durgin (2008). According to Durgin's perspective, numerosity is not a primary visual feature, but a feature that is estimated, similarly to texture density, through stages of spatial filtering, akin to the estimation of ensemble statistics, like average size, as demonstrated by Ariely (2001). To establish that numerosity is a primary feature directly extracted and not indirectly computed, Arrighi et al. (2014) conducted several adaptation studies across modalities and formats (Figure 1.3). Their research revealed that adapting to a given number of sequentially presented tones, for instance, influenced the perceived numerosity of visual flashes and that the effect size is similar to that observed in within-modality adaptation contexts. This observation implies that, at a certain level, the representations of numerosity are not dependent on the modality or format of the stimuli to be enumerated, thus, they do not depend on sensory-specific systems. Previous studies had similarly suggested that both children (Barth et al., 2005) and adults (Barth et al., 2003) exhibit comparable accuracy in cross-modal comparison as they do in visual comparison, suggesting no performance cost associated with numerosity comparison across sensory modalities. These cross-modal findings suggest that the perception of numerosity can be entirely independent of the perception of other sensory features, thus contradicting Durgin (2008) claim.

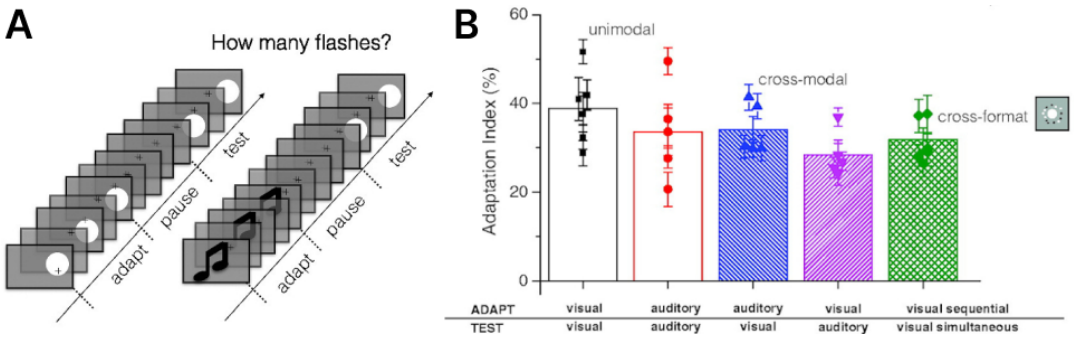



FIGURE 1.3: (A) Example of stimulus for the cross-modal numerosity adaptation. (B) Mean adaptation index for different experimental settings (figure adapted from Anobile et al. (2015))

## 1.2 Animal Neurophysiology

Neurophysiological recordings in animals have revealed single neurons with distinct responses to different numerosities. R. F. Thompson et al. (1970) discovered single neurons in the association cortex of anesthetized cats that seemed to encode numerosity during sequential presentations of visual and auditory items. Neurons encoding numerosity have subsequently also been identified in monkeys through various tasks (Figure 1.4), including movement executions (Sawamura et al., 2002), delayed match-to-sample (Nieder & Miller, 2003, 2004; Nieder et al., 2002), and saccade to target (Roitman et al., 2007). Wagener et al. (2018) more recently demonstrated that in numerically naive crows, neurons in the endbrain encode visual numerosity. Kobylkov et al. (2022) could also find number neurons in the caudal nidopallium (higher associative area with functional similarities to the mammalian prefrontal cortex) in naive 10-day-old domestic chicks. Neurons encoding numerosity have been mainly observed in the intraparietal sulcus (IPS) and subregions of the prefrontal cortex of the macaque brain. However, very few regions were studied, notably the early visual cortex and the occipitotemporal cortex (see Figure 1.4). The tuning functions of these neurons resemble bell-shaped functions with asymmetric profiles; that become symmetric Gaussian tuning functions only after a log-transformation of the number scale (Nieder & Miller, 2003). In addition to the number-selective neurons in the IPS and prefrontal cortex, monkeys' lateral intraparietal area (LIP) also contains number neurons that, contrary to the ones previously mentioned exhibit a monotonic response to numbers (Roitman et al., 2007).

It was proposed that the tuning shape of neurons might not be innate but rather a consequence of training (Roitman et al., 2012). This hypothesis was proven wrong by a later study conducted by Viswanathan and Nieder (2013), who identified number-selective neurons in the intraparietal sulcus and the dorsolateral prefrontal cortex of monkeys that had never undergone number discrimination training. In all of the above studies with animals, other covarying factors such as size and density were individually controlled in different trials, as in many experiments with human subjects.

## 1.3 Human Neuroimaging

Various studies using adaptation paradigms (Cantlon et al., 2006; Piazza et al., 2004, 2007; Tsouli et al., 2021), frequency tagging techniques (Van Rinsveld et al., 2020, 2021), multivariate decoding and representational similarity analysis (Bulthé et al., 2014, 2015; Castaldi et al., 2019; Eger et al., 2009), and population receptive field mapping (Harvey & Dumoulin, 2017a; Harvey et al., 2013) consistently found that the posterior parietal cortex plays a pivotal role in the processing of visual numerosity.

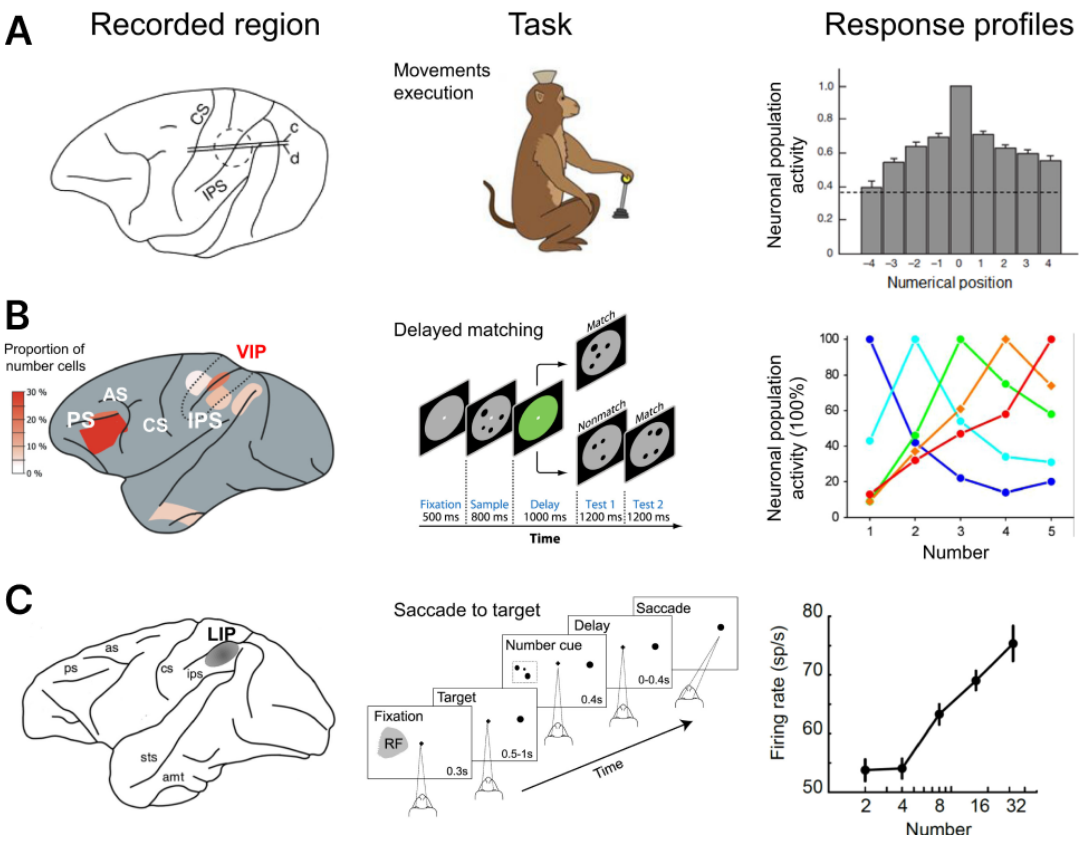



FIGURE 1.4: Neurophysiological evidence of encoding numerosity in monkey's brain with different experimental setting: (A) Sequential movement task, (B) Delayed match-to-sample task, and (C) Non-numerical saccade to target task (figure adapted from Piazza and Eger (2016))

Piazza et al. (2004) were pioneers in using fMRI adaptation to explore visual numerosity representation, as depicted in Figure 1.5 Their study revealed that after a period of adaptation to a specific numerosity with a constant large number of elements (16 or 32, in different blocks) but variable low-level features like size and density, a change in numerosity resulted in a release from adaptation in the bilateral intraparietal cortex. Consistent with Weber's law, the variation in IPS activation reflected the ratio between the adapted and deviant numbers.

Prompted by this initial study, later Eger et al. (2009) focussed their MRI acquisition on parietal regions and employed a multivariate decoding approach to assess the discriminability of sample numerosities in the bilateral intraparietal cortex. Participants were shown sample numbers and instructed to remember them during the task. Other studies with slightly varied task settings (Bulthé et al., 2014, 2015) successfully differentiated between various numerosities using a multivariate decoding method in regions in the parietal and frontal cortex. However, the interpretation of the results from the multivariate decoding approach is not straightforward, as uncertainty persists regarding whether the information crucial for successful decoding in multivariate classification analysis, when collapsing across the non-numerical

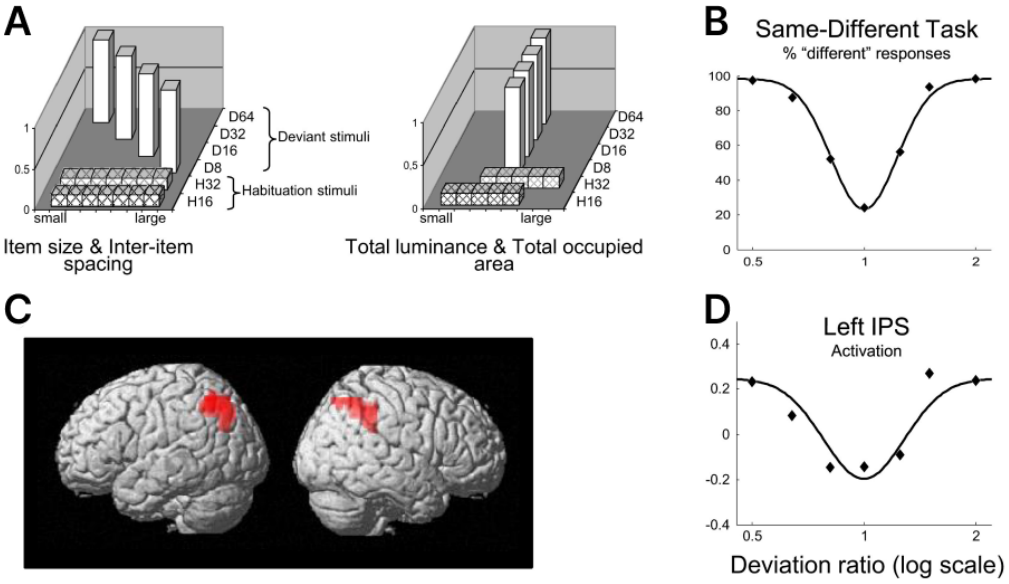



FIGURE 1.5: Using adaptation paradigm Piazza et al. (2004) showed that (A) by presenting deviant stimulus after adaptation stimulus. (B) Gaussian-shape distribution would be produced both in the behavior and (C,D) bilateral mid-posterior parietal cortex (figure adapted from Piazza and Eger (2016))

dimensions of the stimulus set, was exclusively numerical in nature.

The role of the parietal area in representing numbers has been also highlighted by the finding that this region hosts numerotopic maps (Harvey et al., 2013). Harvey et al. (2013) identified a cluster of voxels situated in the posterior superior parietal lobule that displayed a preference for different numbers, and that those clusters were spatially organized reflecting numerotopy (Figure 1.6). Later Harvey and Dumoulin (2017a) reported that these numerotopic maps were present also elsewhere in the brain, and found six such maps across the human association cortex, spanning regions in the parietal, frontal, and occipito-temporal cortex. While their initial studies constrained the number range from 1 to 7, their later research (Cai et al., 2021) extended beyond this range and they found that both small and large numbers are organized within the same topographic maps. To generate the numerosity map, population receptive field (pRF) modeling, as proposed by Dumoulin and Wandell (2008), was employed. This method enables us to predict the cumulative response of populations of cells contained within a single fMRI voxel to various stimuli. However interpreting results from pRF modeling is not without challenges (Gebuis et al., 2014). It is possible that the variance in the fMRI BOLD signal observed in the data can be attributed to a combination of other correlated non-numeric features along with number.

To address the challenges mentioned above, Harvey and Dumoulin (2017b) tried to directly quantify the effect of non-numeric features. However, they primarily focused on the impact of one non-numeric variable at the time, comparing it to that

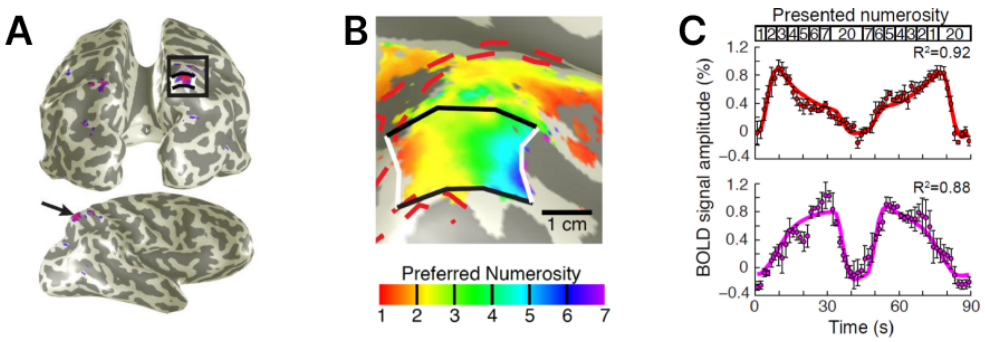



FIGURE 1.6: Harvey et al. (2013) could find numerotopic map in the(A, B) right posterior superior parietal cortex of subjects (C) while they were seeing patterns of dots that increase and decrease in numerosity (figure adapted from Piazza and Eger (2016))

of number, without considering the effects explained by all relevant non-numerical dimensions simultaneously. DeWind et al. (2019) tried to address the issue of correlated non-numeric dimensions by introducing two orthogonal dimensions referred to as "size" and "spacing," [2] (DeWind et al., 2015). However, the study did not consider the perceptual discriminability of these features, as "size" and "spacing", that were introduced by them, were mathematical constructs and did not represent natural, perceptually relevant feature dimensions. Only a recent study by Castaldi et al. (2019) have quantified the effect of numeric and perceptually relevant non-numeric features on the brain activity simultaneously. Castaldi et al. (2019) used a multiple regression approach in order to disentangle brain signals related to numerical and non-numerical quantities and showed that numerosity is represented over and above other visual features in parietal association areas. Moreover, the study revealed that the representation of numbers is not only present in parietal regions but is also present in early visual areas, a finding in line with previous multivariate decoding fMRI study (DeWind et al., 2019; Lasne et al., 2019).

Few studies have investigated the timing of numerosity encoding in the human brain (Fornaciai et al., 2017; Park et al., 2015). The electroencephalogram (EEG) study conducted by Park et al. (2015) was among the pioneering research efforts to explore the temporal representation of numerosity in the adult human brain. Employing a regression-based method with three orthogonal regressors—number, size, and spacing (DeWind et al., 2015)—they analyzed univariate event-related potential (ERP) signals of EEG. The findings revealed that numerosity explained a significant portion of the variance in the ERP very early in time, approximately around 75 ms after stimulus onset. Subsequent research with the same experimental design and analytical approach replicated this outcome, demonstrating that numerosity information is encoded remarkably early, around 90 ms post-stimulus (Fornaciai et al.,

[2] According to their definition, size was obtained through a linear combination of the logarithm of the total surface area and the average item area. Spacing, on the other hand, was derived by a linear combination of the logarithm of the total field area and the inverse of density.

2017). However, as explained earlier, interpreting these results is not straightforward because the two orthogonal dimensions, "size" and "spacing," are mathematical constructs and not perceptually relevant non-numeric features. Consequently, there is a lack of clear information regarding the timing of perceptually defined non-numeric features.

## 1.4 Computational Modeling

Dehaene and Changeux (1993) introduced one of the earliest connectionist models of numerosity perception (Figure 1.7). According to their model, objects are initially mapped onto a location map. Through the process of lateral inhibition, the location map disregards the size of objects, a mechanism known as normalization. In the subsequent summation stage, units aggregate information from the previous layers. This layer acts as an integrator. Finally, the summation units project to numerosity units through central excitation and lateral inhibition connections. The model posits that numerosity detection is innate and present from birth, hardwired and implemented thanks to a pattern of on-center off-surround units. The concept of the location map might be linked to the saliency map in the parietal cortex (Roggeman et al., 2010). Knops et al. (2014) tested this idea using fMRI and computational modeling, exploring the possibility of a flexible system resembling a saliency map underlying an enumeration task. Additionally, Wurm et al. (2021) demonstrated that estimating a small number of objects involves two sequential stages: a location-specific stage followed by a location-invariant stage. This finding aligns with an fMRI study conducted by Eger et al. (2013), which highlighted numerical information encoded in a location-invariant manner in specific areas of the human posterior parietal cortex.

Another computational model of numerosity perception was proposed by Stoianov and Zorzi (2012), who took a different approach to model numerosity encoding by training a Deep Belief Network (DBN) in an unsupervised manner, without providing any explicit information about numerosity of object in the images. They discovered that numerosity selectivity emerges as a high-order statistical feature in the deepest layer of the network, which learns a hierarchical generative model of the sensory input. This modeling approach aligns with Dakin et al. (2011), where they suggested that numerosity is encoded through low-level visual features like contrast or spatial frequency. Additionally, Stoianov and Zorzi's model effectively separates non-numerical features from numerical information (Testolin et al., 2020).

Recently, studies have demonstrated that Convolutional Neural Networks (CNN; Figure 1.7) have the capability to represent numbers (Kim et al., 2021; Nasr et al., 2019). Nasr et al. (2019) found that number-detector units emerged in a CNN trained for visual object recognition. Subsequently, Kim et al. (2021) showed that these number-detector units even emerged in an untrained CNN. Notably, they discovered that these number-selective units were not influenced by non-numeric visual

features, thus suggesting that the extraction of numerosity occurs through a dedicated mechanism in these networks.

Formal models, through providing quantitative description of a system, aid us to understand how the brain processes information. The two key criteria that define a good model are its accuracy and understandability (Kay, 2018). Accuracy refers to how well the model can predict human behavior or brain data at various levels, such as single neurons or populations of neurons. Understandability pertains to how well the scientific community can grasp each component of a model and the relationships between these components and the model's outcomes. To date, to the best of my knowledge, there has been no attempt at investigating how well computational models of numerosity processing align with human population-level neural data. Thus, in the last chapter of my thesis I will focus on investigating the accuracy of one particular model at simulating the human fMRI data that I illustrate in chapter two.

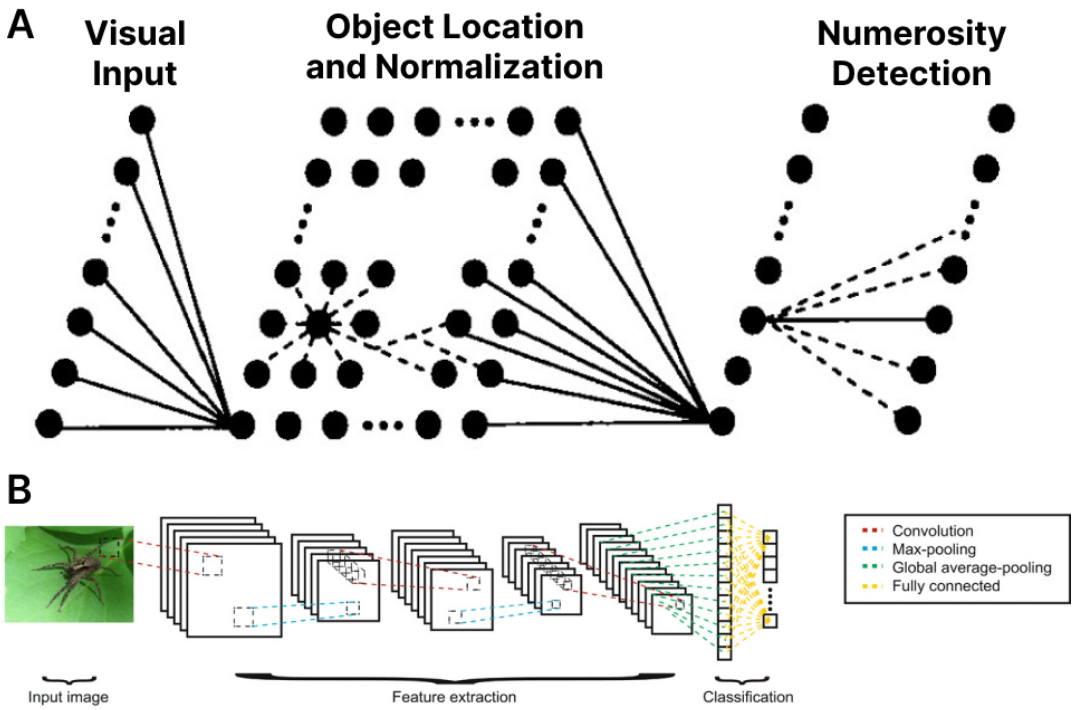



FIGURE 1.7: (A) Structure and function of numerosity detector network (figure adapted from Dehaene and Changeux (1993)). (B) The convolutional neural network architecture employed by Nasr et al. (2019) to demonstrate the network's sensitivity to numbers (figure adapted from Nasr et al. (2019)).

## 1.5 Questions addressed in this thesis

Thus far, I have reviewed various behavioral, neurophysiological, and neuroimaging methods employed in the literature to unveil how numerosity is encoded in humans and animals. In this thesis, I will use the approach employed by Castaldi

et al. (2019) to further investigate the question of how numerosity is encoded in the human brain using different neuroimaging methods. I chose model-based representational similarity analysis because it enables us to concurrently model the effects of both numeric and non-numeric features of stimuli, an issue that is very important to consider.

In particular, in this thesis I will address three different but related questions:

1. The first question pertains to the **localization** of numerical representations in the cortical space. While most neuroimaging studies on number processing have concentrated on the dorsal stream, particularly the parietal region, as a pivotal area for number representation, findings from numerotopic maps (Cai et al., 2021; Harvey & Dumoulin, 2017a) have provided some suggestions that numerosity may be encoded also beyond the classical regions in the parietal and frontal lobes. Consequently, in the **second chapter** of this thesis, I used whole brain functional magnetic resonance imaging (fMRI) to address whether other brain regions, especially along the ventral stream, represent numerical information when subjects are tested with visual sets of dots with the rigorous control over other non-numeric features which lacked in previous research.

2. The second question pertains to the **temporal characterization** of numerical representations in the brain. Indeed, while there has been extensive research in the field of numerosity perception focusing on localization, there have been relatively few studies investigating the time course of the neural response to numerosity. In the **third chapter**, employing the same control over other non-numeric features as used in the first chapter, I use magnetoencephalography (MEG) to examine the temporal unfolding of numerosity representations in the adult human brain. This approach enabled us to tackle the second question: when numeric and non-numeric features of a visual set of dots are separately represented in the brain. Furthermore, I extended my exploration by integrating fMRI data from the previous chapter with MEG data (an approach that is referred to as "fMRI-MEG fusion"), aiming to investigate the spatio-temporal representation of each feature. This approach involved studying the temporal unfolding of the representation of both numeric and non-numeric features across different brain areas.

3. The third question pertains to the question of whether deep neural networks as **computational models** are able to **accurately simulate** how numerical and non numerical information are encoded in the brain. Despite extensive computational research aimed at understanding how numerosity information can be extracted from a visual set of dots, we still do not know the mechanisms through which the brain processes this information. There has been limited effort to compare these models against neural data, although there are exceptions (see Nasr et al., 2019). Thus, in the fourth and final chapter, I employed

the same method used to analyze the fMRI and MEG data in the first and second chapters to examine how CNNs represent numeric and non-numeric features of visual sets of dots. I subsequently compared the results from a CNN against fMRI data from the second chapter to address the third question: determining the extent to which the representations of numerical information in the CNN align with the fMRI data of the human brain.

I will address each of the three questions in the upcoming chapters, each of which has been written independently of the others and will result in three separate scientific publications.

# CHAPTER 2

## Neural codes for visual numerosity in the dorsal and ventral stream of the human brain

# Chapter 2

# Neural codes for visual numerosity in the dorsal and ventral stream of the human brain

## 2.1 Introduction

Single cell recording studies indicate that both in the animal (e.g., Kobylkov et al., 2022; Nieder & Miller, 2003; Wagener et al., 2018) and the human brain (Kutter et al., 2018) there are numerosity-tuned neurons, whose Gaussian tuning functions demonstrate scalar variability, a proportional increase in the standard deviation of numerical estimates with the quantity being estimated. The activity of these functions underlies behavior in numerosity comparison and matching tasks, similarly adhering to Weber's law (Ditz & Nieder, 2016; Gallistel & Gelman, 1992; Merten & Nieder, 2009; Piantadosi & Cantlon, 2017). This approximate and compressed code can also be inferred from the population-level responses to numerosity as measured by fMRI using multivariate pattern analyses, adaptation, or population receptive field mapping approaches (e.g., Eger et al., 2009; Harvey et al., 2013; Piazza et al., 2004). However, because numerosity is necessarily coupled with other visual characteristics of the sets (e.g., more items tend to occupy a larger area, or to be more densely spaced), establishing the degree to which the observed neural response to numerosity is distinct from the response to other visual attributes is not trivial. To address this issue Castaldi et al. (2019) recently analyzed the human BOLD response to numerosity combining multivariate analyses with multiple regression, allowing the estimation of the brain activity evoked by different numerosities once taking into account the effects of other relevant non-numerical variables at the same time (Castaldi et al., 2019). This study however solely focussed on the dorsal stream and demonstrated that numerosity is represented over and above other visual features in retinotopic regions of interest and that, especially when task relevant, it is amplified in parietal areas. In terms of localization, these results were in agreement with several previous observations, including the first single cell recordings in macaques (Nieder & Miller, 2003, 2004) and the first fMRI studies in humans (e.g.,

Ansari & Dhital, 2006; Cantlon et al., 2006; Castelli et al., 2006; Piazza et al., 2004) that suggested that parietal cortex is the key brain region for numerosity processing (for a review see: Eger, 2016; Faye et al., 2019). They were also consistent with the neuropsychological literature that associates deficits in numerosity processing to parietal cortex damage (Lemer et al., 2003; Warrington & James, 1967). Partially biased by these initial observations, several later key fMRI studies on numerosity restricted the brain data acquisition to a limited volume centered on parietal cortex (e.g., Castaldi et al., 2019; Eger et al., 2009) or focussed the analyses on parietal cortex using an ROI approach (e.g, Bulthé et al., 2015; Castaldi et al., 2020). While a few studies looking at the whole brain did sometimes report numerosity-related response outside parietal cortex, both in the frontal and in the occipito-temporal cortex (e.g., Bulthé et al., 2014; Harvey & Dumoulin, 2017a), they mostly tested small numerosities, which some suggest are elaborated by a dedicated mechanism, referred to as "subitizing", potentially different from the one at play with large numerosities (Kutter et al., 2023; Revkin et al., 2008). A notable exception is the recent work of Cai et al. (2021), who employed a population receptive field mapping (pRF) method (Harvey & Dumoulin, 2017a) and found that the numerotopic maps that encode small numerosities in the dorsal and ventral stream also respond to large numerosities. A couple of important limitations from this study, however, call for further confirmations for this stand-alone report:

1. In the study there was no attempt at controlling/analyzing the impact of non-numerical features. In all stimuli total surface area was held constant across numerosities, thus as a consequence dot size decreased and density increased with number. This prevents concluding that the activation to numerosity was solely due to numerosity and not to concurrent changes in other visual features (Gebuis et al., 2014) especially as Harvey et al. (2015) showed numerosity and object size are associated in overlapping topographic maps..

2. The strictly ordered presentation of numerosities (ascending or descending), while ideal for pRF approach, likely created expectations (e.g. Esterman & Yantis, 2009; Puri et al., 2009; Summerfield & De Lange, 2014) and attentional biases (e.g. Ester et al., 2016; Jehee et al., 2011; Lage-Castellanos et al., 2022), the effect of which cannot be readily disentangled from the effect of numerosity itself.

Thus, the question of whether large numerosity is encoded over and above the other visual properties solely along the dorsal stream, or whether it is also represented in other regions of the human brain still remains open. In order to probe this questions, here I adapted the paradigm of Castaldi et al. (2019) and recorded the BOLD signal from the whole brain of subjects looking at sets of different number of dots, and analyzed the data using a combination of model-based representational similarity and dimensionality reduction analyses, both in pre-defined ROIs and across the whole brain.

## 2.2 Methods

### 2.2.1 Participants

Thirty-seven healthy adults (twenty-two females; mean age 21.9 years) with normal or corrected vision participated in the study. The sample size for our study was established based on a prior study by Castaldi et al. (2019), which involved twenty healthy adults. Our chosen sample size is comparable to or even exceeds the typical sample sizes employed in similar experiments in the field (e.g. DeWind et al., 2019; Eger et al., 2009, 2013). The ethics committee of the University of Trento (Italy) approved the study, and all participants gave written informed consent and were reimbursed for their time. Given their excessive head motion (translation, exceeding 3mm in any of the x, y, or z directions, or rotation, surpassing 2 degrees around any of the axes) or poor behavioral performance (accuracy < 65%), data from six participants (four participants for poor behavioral performance and two participants for excessive head motion) were excluded from the final analysis. This led to the final sample of thirty-one subjects (eighteen females; mean age 21.9 years).

### 2.2.2 Stimuli and procedure

Participants were familiarized with the task by practicing 20 trials outside of the MRI before the experiment. During fMRI scanning, arrays of black dots on a mid-gray background were centrally shown to participants. Dots were generated orthogonally varied in number, average item area, and total field area (similar to Castaldi et al., 2019). There were 32 conditions (resulting by crossing 4 numerosities, 4 average item areas, and 2 total field areas): six, ten, seventeen, or twenty-nine dots were presented with varying average item area (0.04, 0.07, 0.12, 0.2 visual square degree) that were made to fit within a small or large total field area (defined by a virtual circle of either about 9 or 13.5 visual degree diameter; Figure 2.1). Numbers and average item areas were chosen based on a previous behavioral study (Castaldi et al., 2018, 2019) to be perceptually equally discriminable. Total field areas were selected to have suitably sparse arrays of dots (1 dot/vd$^2$) to be within the numerosity estimation regime and not within the density estimation regime (Anobile et al., 2013).

In each trial, a set of dots was presented for 500 ms over a wide red fixation cross, and participants were required to estimate their number and simply keep them in memory until, after a variable ISI of 3.5–5.5 s, the next set of dots appeared (Figure 2.1). When the color of the fixation cross changed from red to green, subjects were required to compare the number of dots in the current set (match stimulus) with the previous one and determine whether it was larger or smaller by pressing one of two buttons following the instructions. After a delay of 3.5–5.5 s the following trial started. Match stimuli were designed to be approximately 2 JNDs larger or smaller in numerosity than the previous sample stimulus, based on an average numerosity Weber fraction estimated on an independent group of healthy adults

(Castaldi et al., 2018), while the other dimensions (total field area and average item area) were the same. Match trials occurred approximately 20% of the time.

The experiment consisted of six runs, with two blocks within each run. Each block consisted of thirty-six trials: four match trials and thirty-two sample trials, one for each of the thirty-two conditions (4 number × 4 average item area × 2 total field area). After the third run, in the middle of the experiment, participants' response assignments were switched. There was a brief practice session at the beginning of the experiment and after changing the hand assignment. Each run lasted about 7 minutes.

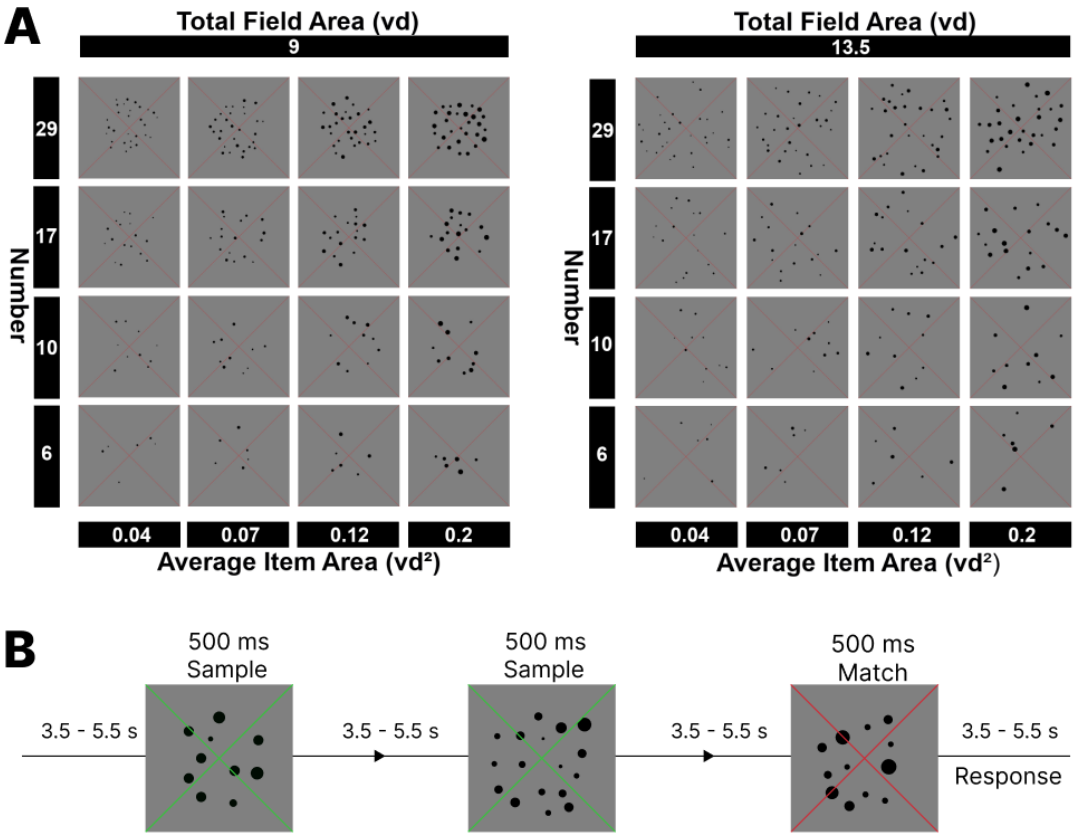



FIGURE 2.1: (A) An illustration of the entire collection of stimulus conditions. The dataset orthogonally varied in number (6, 10, 17, 29), average item area (0.04, 0.07, 0.12, and 0.2 visual square degree), and total field area, enclosed by imaginary circles of 9° and 13.5°. (B) Illustration of the temporal presentation of the trials during scanning. Subjects were instructed to attend to the number of dots and keep the number in mind until the next set of dots were shown. From time to time, the color of the fixation cross changed from red to green. When the color changed, subjects were required to compare the number of dots in the current set, (Match stimulus), with the previous one (Sample stimulus) by pressing a button.

### 2.2.3 MRI recordings and processing

Functional images were acquired at the Center for Mind/Brain Sciences (CIMeC) with a SIEMENS MAGNETOM PRISMA 3T with gradient insert of 80mT/m max and using a SIEMENS Head/Neck 64-channel phased array coil. Visual stimuli

were presented through a mirror system connected to a 42" LCD monitor (MR-compatible, Nordic NeuroLab) positioned at the back of the magnet bore. Functional images were acquired using echo-planar (EPI) T2*-weighted fat-saturation echo-planar image (EPI) volumes with 1.75 mm isotropic voxels using a multi-band sequence (Moeller et al., 2010) (https://www.cmrr.umn.edu/multiband/, multi-band [MB] = 3, GRAPPA acceleration with [IPAT] = 0, partial Fourier [PF] = 7/8, matrix = 120×120, repetition time [TR] = 20.3 ms, echo time [TE] = 31.2 ms, echo spacing [ES] = 0.62 ms, flip angle [FA] = 60°, bandwidth [BW] = 2450 Hz/px, phase-encode direction Anterior » Posterior). In total, 1206 volumes from the six experimental runs made up the functional acquisition. A whole-brain gradient echo B0 map, matched for spatial resolution, was acquired after the functional scans for fieldmap-based correction of susceptibility-induced geometric distortions. T1-weighted anatomical images were acquired at 1 mm isotropic resolution using an MPRAGE sequence (GRAPPA acceleration with [IPAT] = 2, matrix = 176×256, repetition time [TR] = 2530 s, echo time [TE] = 1.69 ms, time of inversion [TI] = 1100 ms, flip angle [FA] = 7°, bandwidth [BW] = 650 Hz/px,). Padding and tape were used to reduce head movement. In their left and right hands, participants each held two response buttons. Stimuli were presented using a custom-written Psychtoolbox 3 (Brainard, 1997) script running on top of MATLAB R2018.

Functional images were preprocessed in MATLAB R2019 using the Statistical Parametric Mapping Software (SPM12, https://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Preprocessing included the following steps: Functional images were slice-time corrected to the middle slice, realignment of each scan to the mean of each run, co-registration of the anatomical scan to the mean functional image, and segmentation of the anatomical image into native space tissue components. No smoothing was applied.

The preprocessed EPI images (in subjects' native space) were high-pass filtered at 128 s and pre-whitened by means of an autoregressive model AR(1). A general linear model (GLM) was used to estimate subject-specific beta weights. For each run, thirty-two regressors of interest were included for each sample stimuli (4 number × 4 average item area × 2 total field area). Regressors for match stimuli, left hand, and right hand were also included. Nuisance regressors were identified with The PhysIO toolbox (Kasper et al., 2017) using six motion parameters and CompCor with five components (Behzadi et al., 2007) and were included in the GLM.

The surface of each subject was generated using Freesurfer 6 (Fischl, 2012) (https://surfer.nmr.mgh.harvard.edu/). The surfaces were then converted to a SUMA standard mesh of 141,000 nodes per hemisphere (Saad et al., 2005) from each participant's anatomical scan using algorithms implemented in the Surfing toolbox (Oosterhof et al., 2011) (https://surfing.sourceforge.net) to produce node-to-node anatomical correspondence across participants' surfaces. For each subject, the parameter estimates (beta weights) for each of the 32 regressors of interest were converted

into a t-statistic and projected on the subject-specific cortical surface using AFNI's *3dVol2Surf* (R. W. Cox, 1996) (https://afni.nimh.nih.gov/) with the "average" mapping algorithm, which roughly represents the value at each vertex of the surface as the average value along a line connecting the smooth white matter and pial surfaces.

### 2.2.4 Univariate fMRI activation

The contrast map of sample stimulus against the implicit baseline was smoothed with a Gaussian 4 mm FWHM filter using the SurfSmooth function with the *HEAT_07* smoothing method (Chung et al., 2005). I then performed a surface-based random-effects group analysis for these maps using a one-sample t-test. The result was then corrected using threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) using Monte Carlo simulations with 10,000 permutations, as implemented in the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016) and projected onto the fsaverage surface for visualization (thresholded at $p < 0.01$, one-tailed).

### 2.2.5 Multivariate Representational Similarity Analysis (RSA)

In order to test if and in which brain regions the representations of numerical and non-numerical features of the stimuli could be disentangled I used representational similarity analysis (Kriegeskorte & Kievit, 2013; Kriegeskorte et al., 2008) enabling the assessment of the simultaneous impact of multiple quantitative dimensions on activity patterns.

In every region, which I approached both using a region of interest approach and a searchlight approach (see below), I used the t-statistics from the first-level analysis to extract the neural representational dissimilarity matrix (RDM) by computing the correlation distance (Pearson correlation) between activation patterns for each pair of conditions.

I subsequently used semipartial correlation analysis (Pearson correlation) to assess the extent to which the fMRI pattern dissimilarity structure could be explained by multiple predictor matrices, reflecting the stimuli's dissimilarity across various significant quantitative dimensions: numerosity, average item area, total field area, total surface area and density. To control for the shared variance between the predictors and avoid overestimating the contribution of each variable because the variance that they share is counted several times, I used semipartial correlation analysis (Abdi, 2007). Therefore, the semipartial correlation between a vectorized neural RDM and the model RDM of interest is the portion of unique variance shared between the neural RDM and the model RDM of interest while partialling out the effect of all other models RDM from the neural RDM. To account for different levels of noise in different brain areas, I estimated the noise ceiling in all ROIs (Nili et al., 2014) and normalized the semipartial correlation coefficients with the corresponding noise ceiling (Al-Tahan & Mohsenzadeh, 2021; Khaligh-Razavi et al., 2018). A

schematic of this process is shown in Figure 2.2. It's important to note that my design orthogonally manipulated numerosity, average item area, and total field area. Consequently, numerosity was correlated with both density and total surface area (correlation between number and density predictor matrix = 0.84; between number and total surface area predictor matrix = 0.36).

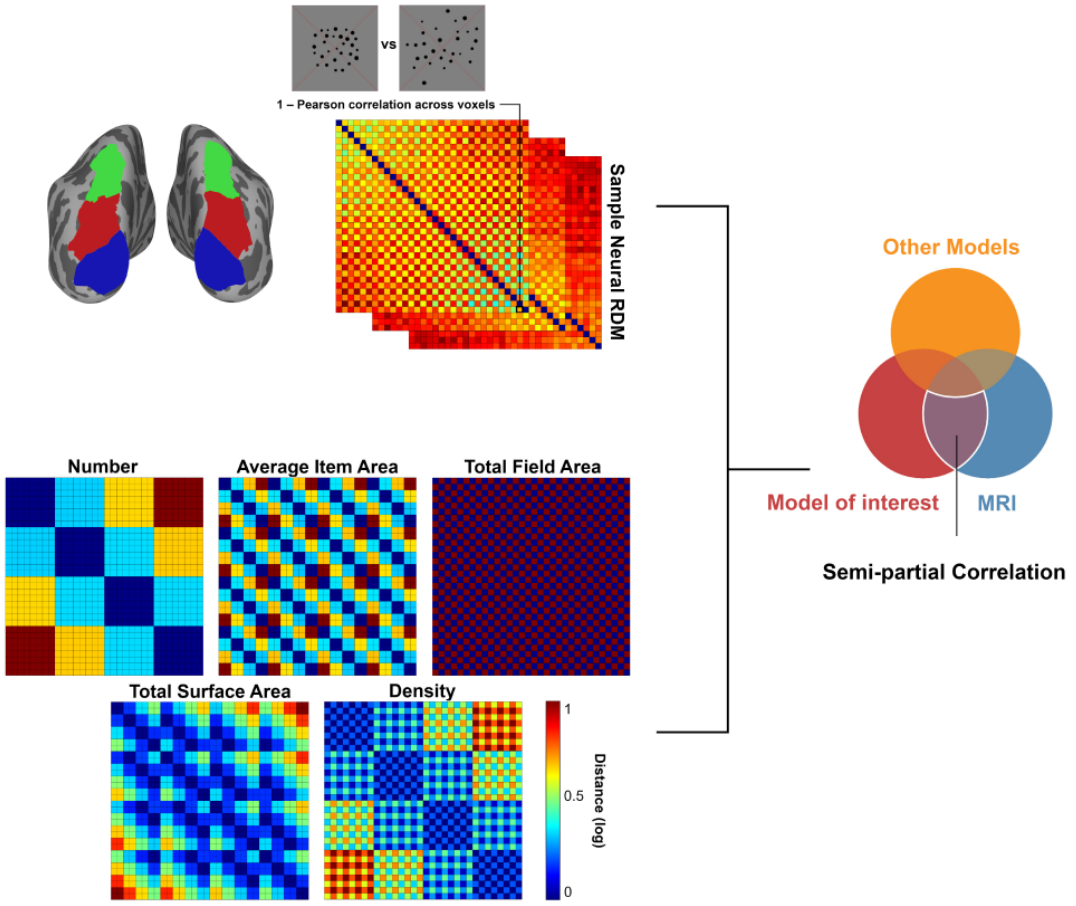



FIGURE 2.2: Neural representational dissimilarity matrices (RDM) obtained from fMRI data were subjected to semipartial correlation analysis. fMRI RDMs were created using 1 – Pearson correlation between the activations of voxels in that region for each pair of images. Five representational dissimilarity matrices, model RDMs, used as predictors in the semi-partial correlation analysis. These matrices represent the logarithmic distance between pairs of stimuli in terms of number, average item area, total field area, total surface area, and density.

### 2.2.6 Surface-based ROI RSA

Following Castaldi et al. (2019), I selected several retinotopic regions of interest (ROIs) along the dorsal stream from a surface-based probabilistic atlas based on visual topography (L. Wang et al., 2014). However, I here extended the analysis to a number of retinotopic regions also of the ventral stream (L. Wang et al., 2014). The selected ROIs (on both hemispheres) along the dorsal stream were: V3AB (merging V3A and V3B), V7, IPS 12 (merging IPS 1 and 2), IPS 345 (merging IPS 3, 4 and 5). These were further merged into three large ROIs that correspond to intermediate

(V3A, V3B, and V7, also known as IPS0), and higher-level (IPS1 to IPS5) areas. The selected ROIs from the ventral stream were hV4, VO1, VO2, PHC1, PHC2. As for the dorsal regions, the ventral ROIs were further merged into three large ROIs: intermediate (VO1 and VO2), and higher-level (PHC1 and PHC2) areas. The ROIs in early visual areas included V1, V2, and V3, which were subsequently merged into a larger early ROI referred to as V13. I chose the 600 most active vertices (in the contrast "all sample stimuli > baseline") in all ROIs to equate the number of vertices in each ROIs to ease comparisons between ROIs (Mitchell et al., 2004). I choose the vertices from each individual ROIs and large ROIs separately. The choice of 600 vertices was made as a compromise between the maximum number of vertices inside all ROIs (661 vertices) and the use of a significantly smaller number of vertices (such as 100 vertices). Selecting the latter, smaller number of vertices, might have resulted in artificially elevated selectivity for the desired contrast, potentially masking the genuine patterns of interest. It is important to note, though, that when I selected all vertices across each ROI the results remained very similar. For the ROI-based RSA I used unsmoothed data but the results remained remarkably consistent when I smoothed data (De Beeck, 2010) with the same method I described previously. ROI-based RSA was implemented using the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016) and custom-written code in MATLAB R2019 (The MathWorks, Inc., Natick, MA). I used one-sample t-tests against zero across subjects to test the statistical significance of correlation coefficients for each feature and ROI. I analyzed the effects of ROI and features with repeated measures analysis of variance (ANOVA).

### 2.2.7 Surface-based searchlight RSA

To find how numerical and non-numerical quantities are represented across the whole cortical surface, RSA was performed using a surface-based searchlight approach (Oosterhof et al., 2011). Surface-based searchlight RSA was implemented using the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016), the Surfing toolbox (Oosterhof et al., 2011), and custom-written code in MATLAB R2019 (The MathWorks, Inc., Natick, MA). Each participant's entire t-statistics brain map was smoothed with the same method I used to smooth the contrast map and underwent a searchlight (radius 6 mm) procedure. A neural RDM was constructed using Pearson's correlation. Similar to the ROI-based RSA, the semipartial correlation between the neural RDM and model RDMs were calculated and then mapped on the brain.

To identify vertices in which the Fisher-transformed semipartial correlation resulting from searchlight RSA was significantly above zero, a one-sample t-test was used across subjects. The result was then corrected using threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) using Monte Carlo simulations with 10,000 permutations implemented in the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016). The resulting statistical map was thresholded at $p < 0.01$ (one-tailed) and was projected on the fsaverage surface for visualization.

### 2.2.8 Multidimensional Scaling (MDS)

Finally, to explore the latent similarity structure of the neural RDMs, indexing the neural representational geometry of the stimuli space, both across and within ROIs, I used multidimensional scaling (MDS; Kruskal, 1964) using the MATLAB function *cmdscale* and visualized the first two dimensions of the MDS output. These techniques organize the stimuli on a two-dimensional plot in a manner where the distances between them mirror the differences in the response patterns they generated. As a result, stimuli positioned closer together in these depictions triggered more akin response patterns (Nili et al., 2014). First, in order to compare the similarity of the neural representation of the stimulus space across ROIs I vectorized the 32 x 32 neural RDM reflecting the correlation across conditions within each of the 12 ROIs and then constructed a 12 × 12 RDM across ROI. Then, to further explore the neural representational geometry in each stream, I applied the MDS on the group-average RDM across participants for each ROI. The group-average RDM is computed by averaging the RDMs for all thirty-one subjects, which are more precise and less noise-prone in comparison to the RDMs obtained from individual subjects (Nili et al., 2014). I also extended these analyses beyond the retinotopic regions including three additional clusters resulting from the whole brain searchlight map of regions encoding numerosity. To isolate these clusters I used group-constrained subject-specific (GCSS) analyses (Fedorenko et al., 2010; Julian et al., 2012), using custom-written MATLAB R2019 code (The MathWorks, Inc., Natick, MA) developed by Scott and Perrachione (2019) (available at https://github.com/tlscott/make_parcels). The identification of these clusters involved a four-step process, as depicted in Figure 2.3: Initially, the Fisher-transformed correlation values from the number searchlight maps were converted into z-scores. Subsequently, these z-scores were thresholded at $p < 0.01$ and then binarized. Secondly, a probability map was generated by overlaying all binary maps. This resulting map was smoothed using a gaussian kernel of 6 mm FWHM, and vertices with contributions from fewer than ten subjects were discarded. Thirdly, the watershed algorithm, as implemented in the SPM-SS toolbox (Nieto-Castañón & Fedorenko, 2012) (https://www.nitrc.org/projects/spm_ss), was employed to locate local maxima in the probability map. Clusters were defined around these local maxima by filling in volumes surrounding them. Initially, vertices with local maxima were labeled, and this labeling was iteratively extended to neighboring vertices until a local minimum or a zero-valued vertex was encountered. The resulting parcels represent regions where multiple subjects exhibited suprathreshold activity, without the requirement that this activity occur in the exact same vertex across participants. Finally, the number of subjects contributing to each parcel was calculated, and parcels where more than 80% of subjects contributed were selected as the final parcels. The advantage of employing this method, as opposed to manual selection based on group mean, lies in its ability to analyze nearby voxels from different subjects without the necessity

of these vertices overlapping precisely in a common space. Additionally, this approach safeguards against the potential bias introduced by a few subjects exhibiting exceptionally high vertex activity during the selection process (Scott & Perrachione, 2019).

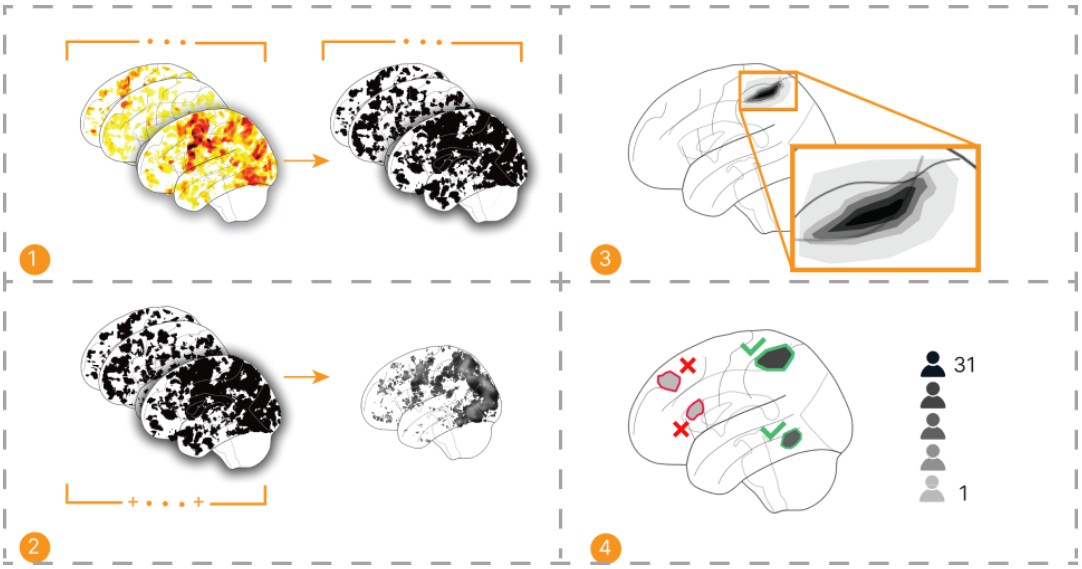



FIGURE 2.3: I employ an algorithmic process to extract parcels in four steps: 1. **Binarization**: Each subject's searchlight map of number (thresholded at p<0.01) is binarized. 2. **Probability Map Creation**: The binary maps are overlaid to generate a probability map, which is subsequently smoothed. 3. **Watershed Algorithm**: The watershed algorithm is applied to identify local maxima, around which parcels are created. 4. **Filtering Criteria**: Parcels are retained if they encompass significant vertices from more than 80% of the subjects.

## 2.3 Results

Thirty-one healthy adult volunteers completed a numerosity comparison task on arrays of dots orthogonally varying in numbers of items (6, 10, 17, 29), average item areas (0.04, 0.07, 0.12, 0.2 visual squares degree), and total field areas (9 or 13.5 visual degree diameter) while being scanned in a 3T MRI. Participants had to keep the number of dots from the sample stimulus in their memory to compare it with an occasionally presented subsequent match stimulus (indicated by a change in color of the fixation cross). When presented with the match stimulus participants were required to decide whether it was more or less numerous than the previous sample stimulus and deliver their response by pressing a button with their right or left hand.

### 2.3.1 Behavioral performance and univariate fMRI activation results

Participants' performance was overall high (Mean=82.14%, SD=6.83%, Range=66.66% – 93.75%), indicating that they were attentive.

Surface-based random-effects group analysis for sample stimulus against the implicit baseline, where participants fixating on a central cross and attended to the

number of dot, revealed activation in a wide set of regions extending dorsally from early visual to parietal up to the postcentral gyrus and the precentral sulcus in the frontal cortex; Ventrally it included medial and lateral inferior occipito-temporal areas (Figure 2.4, thresholded at $p < 0.01$, TFCE corrected).

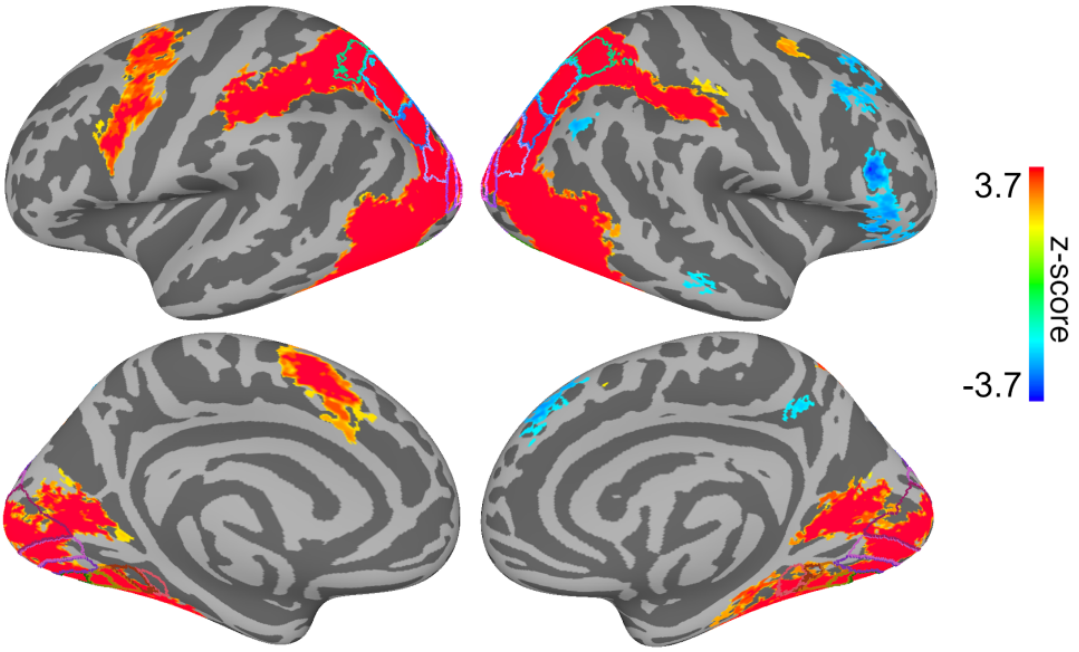



FIGURE 2.4: Results obtained from the surface-based group analysis (n = 31). The maps show the activation elicited for all sample stimuli contrasted against the implicit baseline. Activation maps are thresholded at $p < 0.01$, TFCE corrected, and displayed on Freesurfer's fsaverage surface with colored outlines identifying ROIs along the early visual (V1, V2, V3), the dorsal (V3AB, V7, IPS12, IPS345), and the ventral stream (hV4, VO1, VO2, PHC1, PHC2) from a surface-based probabilistic atlas (L. Wang et al., 2014).

### 2.3.2 Surface-based ROI representational similarity analysis

To unravel the contribution of both numerical and non-numerical features of the stimuli to the distributed activity patterns of the BOLD signal, I performed representational similarity analysis (Kriegeskorte et al., 2008). Figure 2.5 shows the results from semipartial correlation between the neural and model RDMs across different regions in the dorsal and ventral stream, respectively. Semipartial correlation ensures that the resulting coefficients reflect the unique variance explained by each model while partialling out the effect of all other models. The results indicate that the variance in brain activation patterns is significantly explained by number over and above all other non-numerical features, in almost all regions starting already in early visual areas (except V3 and PCH2, all $p<0.01$), and reaching its highest explanatory power in higher-level regions: V7-IPS along the dorsal stream, and VO1-VO2 along the ventral stream. The opposite pattern of results was true for most non-numerical features: total field area explained independent portions of variance in all regions, but maximally in V1-V3 and less so in higher-level regions of both

streams; total surface area was significant only up to V3AB in the dorsal, and hV4 in the ventral stream; density also was significant in V1-V3 and hV4, then stopped to explain significant portion of the variance but regain some effect in IPS1-5. Finally, the average item area was never significant, neither in the dorsal nor in the ventral stream.

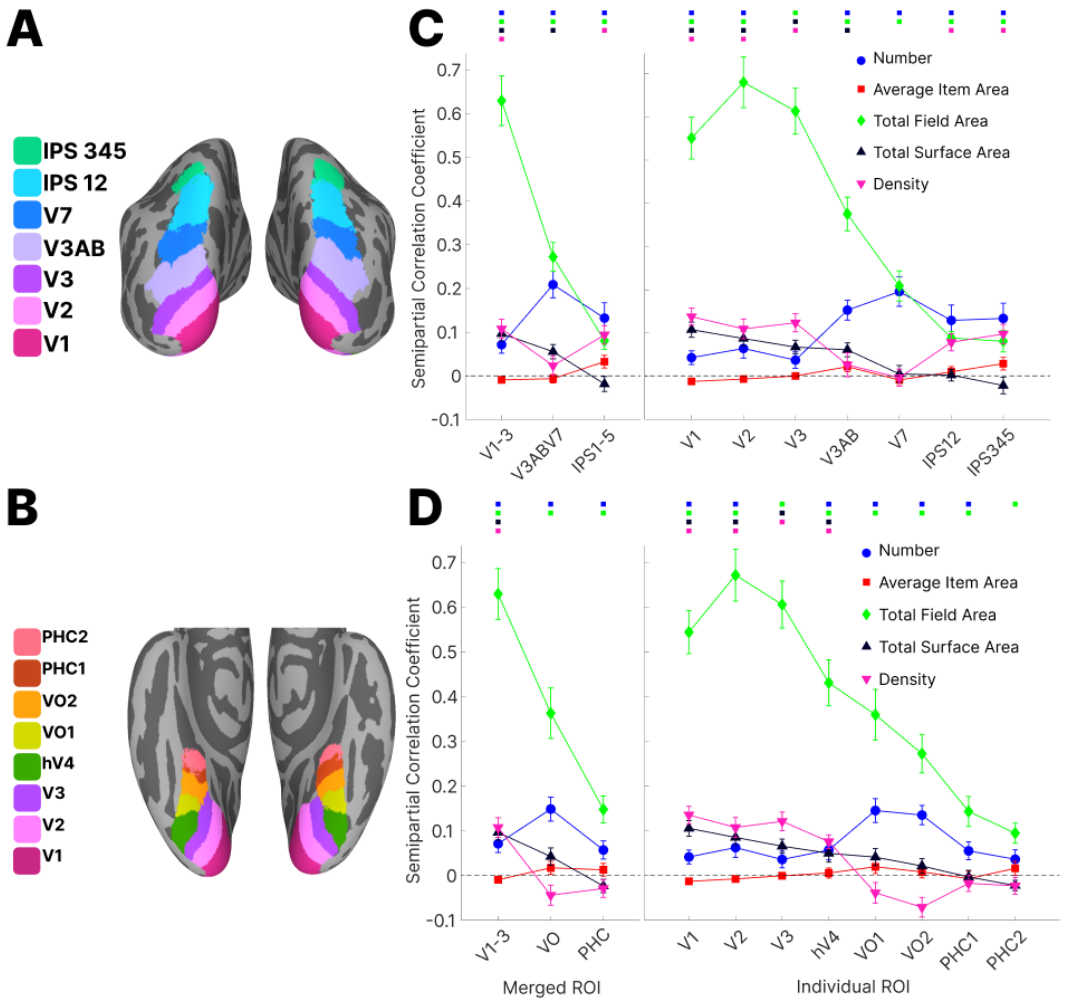



FIGURE 2.5: Color-coded ROIs for (A) dorsal and (B) ventral stream defined by the probabilistic atlas (L. Wang et al., 2014) are displayed on Freesurfer's *fsaverage* inflated surface. (C) Semipartial correlation coefficient obtained from the representational similarity analysis for number, average item area, total field area, total surface area and density from predefined dorsal retinotopic ROIs. Number is represented along other visual features in almost all regions and is amplified in association areas (from V3AB to IPS). Data points show mean semipartial correlation coefficient across subjects (n = 31) ± standard error of the mean (SEM). The coloured points above the figure indicate where the effect significantly exceeds zero ($p<0.01$). (D) Semipartial correlation coefficient obtained from the representational similarity analysis for number, average item area, total field area, total surface area and density from predefined retinotopic ventral ROIs. Within the ventral stream, number is solely represented in intermediate areas (from VO1 to PHC1). Data points show mean semipartial correlation coefficient across subjects (n = 31) ± standard error of the mean (SEM). The coloured points above the figure indicate where the effect significantly exceeds zero ($p<0.01$). All the semipartial correlation coefficients were normalized according to the noise ceiling of their corresponding regions, as detailed in the supplementary materials.

To statistically support the differential impact of the different features across the ROIs, I analyzed the semipartial correlation coefficients with four repeated measures ANOVAs (see below) with ROIs and features as factors. ROIs were considered both aggregated in three big regions, and separately for each of the individual regions, for the ventral and the dorsal stream separately. The significant two-way interaction between ROIs and features that I observed on all the four ANOVAs confirmed that the semipartial correlation coefficient estimated for the different features were different across ROIs in both the dorsal stream hierarchy (for the three large regions: $F(4.30,128.95)=36.680$, $p<0.001$; for the individual regions: $F(8.10, 243.08)=32.404$, $p<0.001$) and in the ventral stream hierarchy (for the three large regions: $F(3.56, 106.91)=29.330$, $p<0.001$; for the individual regions: $F(9.56, 286.82)=27.513$, $p<0.001$).

I then performed five one-way repeated measures ANOVAs on each feature across the ROIs. They revealed that all features were encoded differently across ROIs along the dorsal stream hierarchy (all effects of ROIs for all features $p<0.01$, for both the aggregated and the individual ROIs), and for all features but average item area along the ventral stream (all $p<0.0$, average item area: three large ventral regions: $F(1.91,57.40)=2.304$, $p=0.111$; eight individual ventral regions: $F(5.45,163.43)=1.362$, $p=0.238$ ).

Finally, one-way repeated measures ANOVA on both large ROIs and individual ROIs showed a main effect of features (all $p<0.01$).

### 2.3.3 Surface-based searchlight representational similarity analysis

Besides the region of interest analyses, I sought to investigate how patterns of activity across the cortical surface are captured by the models. To do this, I carried out surface-based searchlight RSA, which revealed that number explained variance in a number of regions widely spread across the cortical surface (Figure 2.6) that encompass but also extend well beyond the retinotopic ROIs. These regions include the mid and anterior parietal cortex, both superior, inferior, and anterior to the intraparietal sulcus, the parieto-occipital cortex, the precentral gyrus and superior frontal sulcus. Along the ventral stream, the regions isolated through the searchlight that were not already part of the previously considered retinotopic ROI areas extended laterally in the mid inferior temporal lobe. These results highlight that numerosity is encoded not only in topographically organized visual areas, but also in a large set of non-retinotopic parietal, temporal, as well as frontal regions. On the contrary, all non-numerical features (apart from average item area, for which I could not find any regions representing it) explained variance mainly in the early visual cortex (Figure 2.6).

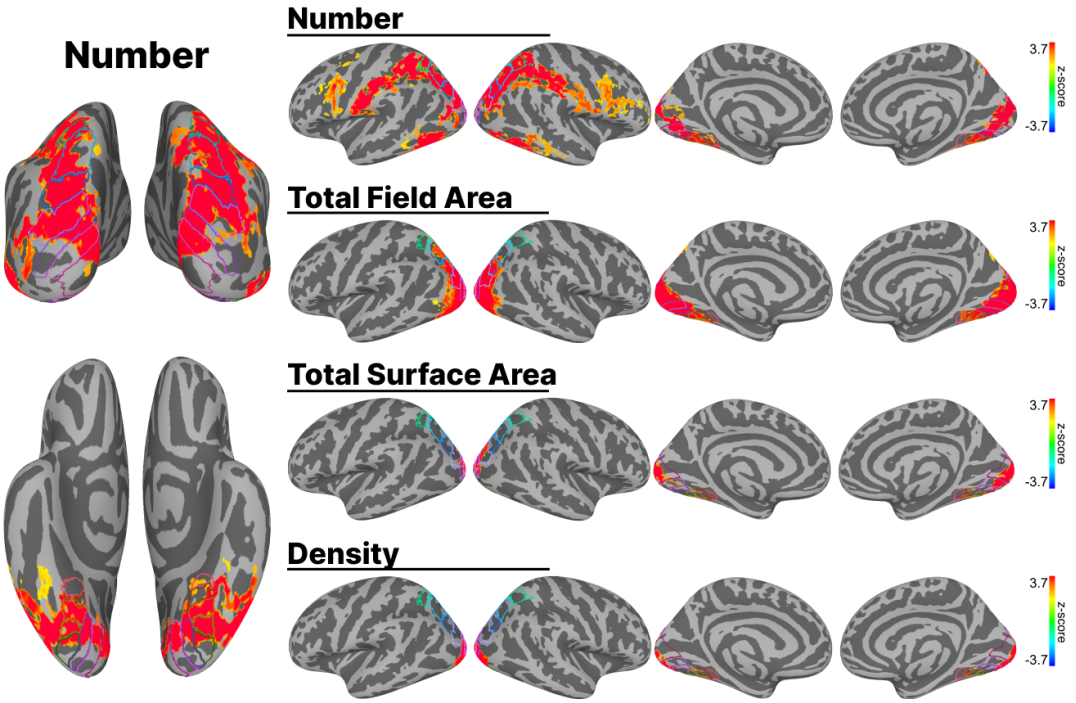



FIGURE 2.6: Surface-based searchlight representational similarity analysis searchlight results obtained from the surface-based group analysis (n = 31). The maps show how patterns of activity across the cortical surface are captured by each model of interest while partialling out the effect of other models. Activation maps are thresholded at $p < 0.01$, TFCE corrected, and displayed on Freesurfer's fsaverage surface with colored outlines identifying ROIs along the early visual (V1, V2, V3), the dorsal (V3AB, V7, IPS12, IPS345), and the ventral stream (hV4, VO1, VO2, PHC1, PHC2) from a surface-based probabilistic atlas (L. Wang et al., 2014).

### 2.3.4 Visualization of similarity structure with multidimensional scaling

To further explore whether the neural representational geometry of the stimuli space was similar or differed across the two streams I used MDS to obtain a low-dimensional representation of the neural similarity structure, and I did it both across ROIs and within ROIs.

The across ROIs MDS (Figure 2.7) suggested the presence of three well distinct clusters: one including early visual regions, the other all ventral regions, and the third one all the dorsal regions, suggesting that the three groups of regions represent the stimuli space differently. This result thus prompted us to further investigate the neural representational geometry within each ROI (see Figure 2.7). In the early visual areas the results indicate a rank-order representation of numerosity along both dimensions, as well as a clear separation between stimuli with a large total field area and with those with a small total field area. While this pattern remains approximately constant along the ventral stream ROIs up to VO2, it changes as I get to the higher ROIs along the dorsal stream (IPS12 and IPS345): here the separation

based on total field area decreases and a curved pattern around the midpoints representing number emerges. This curved pattern is not evident in the ventral stream areas VO1 and VO2. Interestingly, the RSA analysis indicated an equally strong significant representation of numerosity in VO1, VO2, and V7, IPS12, and IPS345. These results highlight that even though numerosity information is equally present in associative retinotopic areas of both streams, the underlying geometry seems different across streams in retinotopic regions. Importantly, however, the observed distinction diminishes when examining regions beyond the retinotopic ROIs, where the method used to consistently select activated parcels revealed three clusters of numerosity-responsive vertices: two in the dorsal stream overlapped with Harvey and Dumoulin's (2017) "NPC1" and "NPC2" and the ventral parcel as "NTO." NPC1 and NPC2 located in and around the postcentral sulcus, while NTO was situated in the lateral occipito-temporal cortex, akin to the regions described by Harvey and Dumoulin (2017a) (Figure 2.8). Specifically, I identified three parcels in the lateral occipito-temporal cortex with the following MNI coordinates in the right hemisphere (46, -71, -6; 46, -67, 3; 47, -60, -6) and left hemisphere (-46, -62, -6; -48, -79, -11; -44, -61, -13). These coordinates appear to be in close proximity to the region reported by Harvey and Dumoulin (2017a) (right hemisphere: 44, -75, -4; left hemisphere: -42, -77, 3).

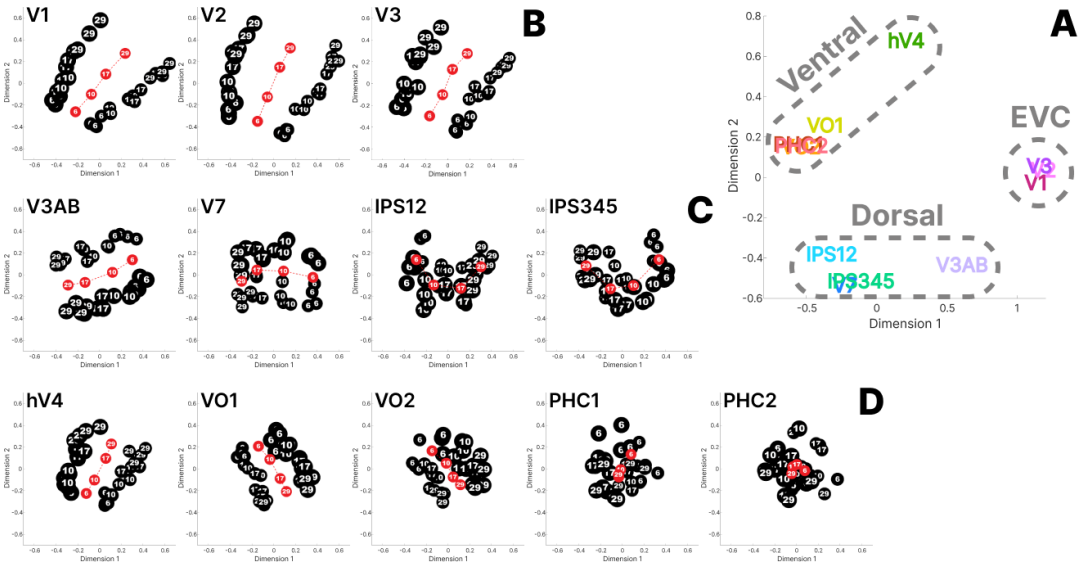



FIGURE 2.7: (A) Multidimensional scaling (MDS) reveals similarity of the representational structures between ROIs in a two-dimensional space. Here the proximity between any two ROIs indicates how similar their representation of the stimuli is. Three clusters of regions become apparent. MDS reveals representational similarities between stimuli in a two-dimensional space for regions in the (B) EVC, (C) dorsal regions, and (D) ventral regions. The black circles represent the 32 stimuli. They are labeled according to their numerosity (6, 10, 17, 29), and scaled in size to represent stimuli with small total field area (small circles) and larger total field area (large circles). The red circles indicate the average coordinates of each number.

All three regions, including the ventral area NTO and the parietal areas NPC1 and

NPC2, exhibited a similar curved structure in response to numerosity (Figure 2.8), indicating a similar representational pattern.

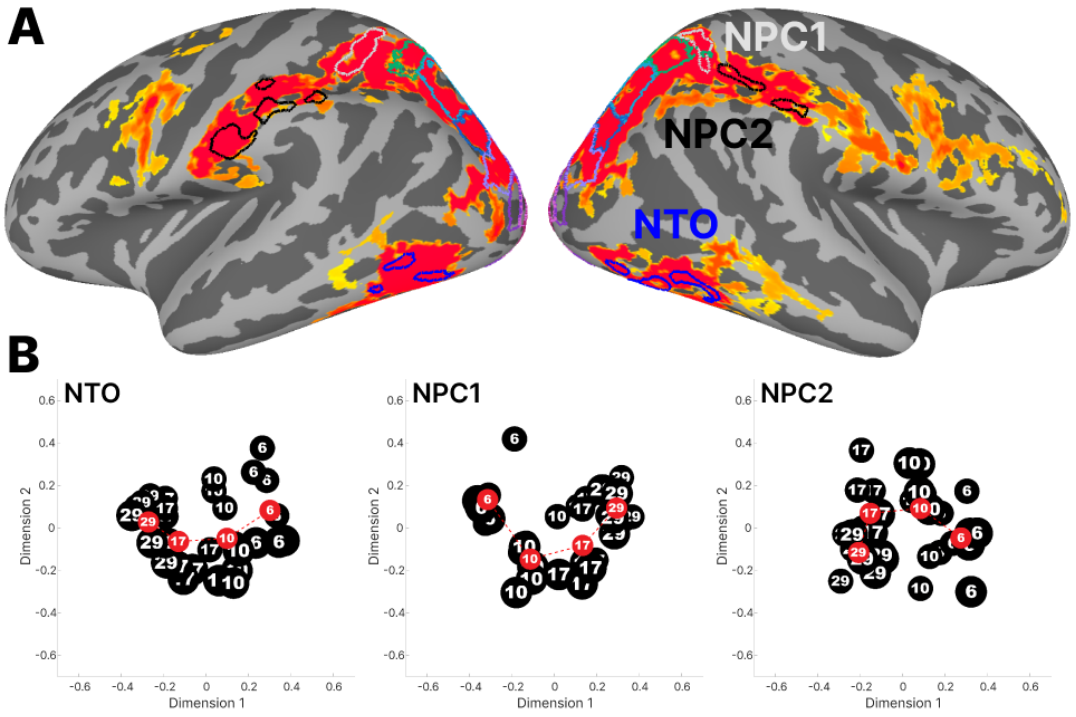



FIGURE 2.8: (A) ROIs (NTO, NPC1, NPC2) chosen using GCSS analysis from outside the retinotopic regions of interest, as defined in the probabilistic atlas by L. Wang et al. (2014), were visualized on the number activation map derived from the searchlight analysis. (B) Multidimensional scaling (MDS) reveals representational similarities between stimuli in a two-dimensional space for regions in the dorsal (NPC1, and NPC2) and the (C) ventral (NTO) stream.

## 2.4 Discussion

The current study aimed to identify neural correlates underlying representation of numeric and non-numeric features of non-symbolic numerical stimuli. In contrast to the existing body of studies that predominantly relied on univariate fMRI techniques or MVPA decoding to address this issue, similar to Castaldi et al. (2019), I employed a different approach called representational similarity analysis (RSA) to simultaneously investigate the impact of numeric and non-numeric features on the blood-oxygen-level-dependent (BOLD) signal. In my study, I extended the work of Castaldi et al. (2019) by applying RSA to regions beyond the dorsal stream, including various regions within the ventral stream hierarchy. This allowed us to investigate the representation of numeric and non-numeric features in a broader network of brain regions involved in visual processing. Moreover, I employed a searchlight approach to examine how patterns of brain activity across the cortical surface were captured by the models. This analysis involved systematically evaluating small spherical regions across the cortical surface to understand how numeric

and non-numeric features influenced neural activity locally throughout the brain.

**Representations of non-symbolic numbers are not restricted to only the dorsal stream**

Similar to Castaldi et al. (2019) the results revealed that the variability observed in activation patterns within the regions along the dorsal stream can be attributed significantly to numerosity when participants attended to numbers. This effect was observed in early visual areas (V1 – V3), and the effect was more pronounced in regions spanning from V7 (a higher visual area) to the intraparietal sulcus (IPS). However, my results were not limited to the regions in the dorsal stream. I could also identify regions in the ventral stream which are sensitive to the number of objects.

At first glance, the obtained results may not appear particularly surprising, considering the accumulating evidence using ECoG (Daitch et al., 2016; Kutter et al., 2022; Pinheiro-Chagas et al., 2018; Shum et al., 2013) and fMRI (Amalric & Dehaene, 2016; Grotheer et al., 2018; Kersey et al., 2019; Yeo et al., 2020) pointing towards the existence of a specialized brain region known as the number form area (NFA). This region is located in the inferior temporal gyrus (ITG) and has been associated with the processing of Arabic numerals (for a review see: Yeo et al., 2017). The previous absence of evidence regarding a consistent localization of the purported number form area (NFA) or visual number form area (VNFA) using fMRI has been attributed to a potential dropout of the fMRI signal in this region due to the proximity of the NFA to the signal-dropout zone generated by the nearby auditory canal and venous sinus (Shum et al., 2013). Nevertheless, recent findings from Grotheer et al. (2018) indicate that it is feasible to examine brain activity within this specific region of the inferior temporal area using fMRI. The aforementioned studies provide compelling evidence for the involvement of the inferior temporal gyrus (ITG) in number processing, none of the previously mentioned studies used non-symbolic representations of numbers, as I did in my own investigation. Instead, some of them employed Arabic representations of digits (Daitch et al., 2016; Grotheer et al., 2018; Pinheiro-Chagas et al., 2018; Shum et al., 2013; Yeo et al., 2020) while others employed arithmetic or math-related tasks (Amalric & Dehaene, 2016; Daitch et al., 2016; Grotheer et al., 2018; Kutter et al., 2022; Pinheiro-Chagas et al., 2018; Shum et al., 2013). Considering the hypothesis that the foundation of symbolic counting and arithmetic in humans can be traced back to non-symbolic representations of numbers (Dehaene & Cohen, 2007; Nieder, 2020; Piazza, 2010) offers a potential explanation for my findings. For instance, in a study by Cai et al. (2023), it was observed that there are overlapping yet distinct tuned neural populations in the ventral temporal-occipital cortex that respond to both symbolic and nonsymbolic numbers. The researchers interpreted these findings as evidence supporting the connection between nonsymbolic and symbolic numerical processing. However, it is important to note that alternative hypotheses argue

that approximate number representations are insufficient for establishing the conceptual content required for integers (Carey & Barner, 2019). For example, Lyons et al. (2014) conducted a study where they directly correlated the activation patterns for symbolic (words, digits) and non-symbolic (visual sets) numbers in the intraparietal sulcus. However, their findings did not provide evidence of a shared representation between these two types of numbers. In my study, I focused exclusively on non-symbolic representations of numbers and did not assess participants with symbolic representations. Consequently, my findings do not provide evidence either supporting or refuting the hypothesis that non-symbolic representations of numbers serve as the foundation for the conceptual content of integers.

Earlier investigations employing whole-brain functional imaging failed to detect evidence of numerosity representation beyond the established network of areas in the frontal and parietal cortices (e.g. Piazza et al., 2004, 2007). This limitation may be attributed to their restricted power, likely stemming from the use of passive or minimally demanding tasks. While more recent studies could identify numerosity representation in the early visual areas (Bulthé et al., 2014, 2015; DeWind et al., 2019) and temporal regions of cortex (Bulthé et al., 2014, 2015), one should interpret these results with caution, as discussed before, because it remains uncertain whether the information crucial for successful decoding in multivariate classification analysis, when collapsing across the non-numerical dimensions of the stimulus set, was exclusively numerical in nature. Additionally, it is worth noting that in the studies conducted by Bulthé et al. (2014, 2015) the range of stimuli employed did not entirely fall within the estimation range, some fell within the subitizing range.

Additional lines of inquiry provide evidence that numerosity representation extends beyond the areas in the frontal and parietal cortices comes from population receptive field (pRF) studies (Cai et al., 2021; Harvey & Dumoulin, 2017a; Harvey et al., 2013; Paul et al., 2022). In these studies, a network of topographic numerosity maps was identified across the brain, encoding numerosity. These Numerotopic maps were categorized as numerosity temporo-occipital (NTO), numerosity parieto-occipital sulcus (NPO), numerosity postcentral sulcus (NPC1, NPC2, NPC3), numerosity frontal sulci (NF). These findings suggest that numerosity representation extends beyond the frontal and parietal cortices. However, it's worth noting that these pRF studies did not encompass the entire brain, and they excluded anterior frontal and temporal lobes from their analyses. Only Cai et al. (2021) conducted full-brain-coverage imaging, but they did not find evidence of numerosity representation beyond the regions they had previously identified. Moreover, this study faces a couple of important limitations. First, in all stimuli, the total surface area was held constant across numerosities, resulting in a decrease in dot size and an increase in density with number. Although they argued that the observed response solely reflects numerosity in their study, as it has been represented in the same numerotopic

maps in their previous studies where they investigated numbers within the subitizing range without any correlation between number and both density and dot size (Harvey & Dumoulin, 2017a; Harvey et al., 2013), this explanation may not be satisfactory. Previously, they have demonstrated an association between numerosity and object size in overlapping topographic maps (Harvey et al., 2015). Therefore, it is also plausible that the reported maps reflect dot size as it decreased with numerosity. Furthermore, in all of their studies, they did not control for the effect of non-numerical features alongside numerosity on the variance explained by the BOLD signal simultaneously, a step we adhered to in our study by employing semi-partial correlation. Therefore, it remains plausible that the numerotopic maps reflect the weighted combination of responses to non-numerical features, from which numerosity can be indirectly computed (Gebuis et al., 2014). Second, to minimize the effects of expectation and attention, we employed a randomized event-related design, whereas they presented stimuli in a strictly ordered manner (either ascending or descending). This approach can introduce biases whose effects cannot be easily separated from the sensory representation of the stimuli themselves. Research has shown that expectations about specific categories can lead to an increase in BOLD signals within category-selective regions, even in the absence of stimuli. For instance, presenting the word 'FACE,' which predicts the subsequent presentation of face stimuli, elicits a higher BOLD signal in the fusiform area (Puri et al., 2009). Similarly, face stimuli gradually emerging from noise result in a larger BOLD signal in the fusiform area (Esterman & Yantis, 2009). Attention can also amplify stimulus-driven responses in the brain (e.g. Castaldi et al., 2019; Ester et al., 2016; Jehee et al., 2011). Recently, Lage-Castellanos et al. (2022) demonstrated that attending to a specific sound feature can induce receptive field changes throughout the auditory cortex. Attention can also modulate population receptive fields when the stimuli are visual sets of dots through both faciliatory and suppressive components (Cai et al., 2022) by a center–surround configuration (Zuiderbaan et al., 2012). In an experiment, Cai et al. (2022) presented participants with both black and white sets of dots, instructing them to focus on either the black or white dots. They proposed that neural responses are a combination of enhanced responses to the numerosity of the attended set and suppressive responses to the numerosity of the unattended set. When stimuli are presented in a strictly ascending or descending order, the anticipation for a certain numerosity may lead to enhanced responses to anticipated numbers and suppression of responses to numerosities deemed improbable to be presented. Therefore, the alternative explanation suggesting that non-numeric features confounding numerosity tuning might more adequately explain the absence of selective tuning for digits, except in the NTO map (Cai et al., 2023), as digits and their low-level features are not confounded in the same way as numerosity.

In addition to fMRI studies, computational modeling using Convolutional Neural Networks (CNNs) has provided valuable support for the hypothesis that number representation is not confined solely to the regions in the dorsal stream. CNNs

have demonstrated their ability to capture brain responses in the inferior temporal cortex (IT), a crucial region for object recognition in both humans and monkeys (Khaligh-Razavi & Kriegeskorte, 2014). By training CNNs on object classification tasks, researchers have successfully accounted for the neural activity observed in the IT cortex. Recent studies have extended the application of CNNs to investigate the representation of numerosity. It has been revealed that number-detecting units, similar to the specialized neurons recorded in the prefrontal and parietal cortex of monkeys, emerge in the final layer of a CNN trained for visual object recognition (Nasr et al., 2019). Surprisingly, even untrained CNNs have shown the emergence of these number-detecting units (Kim et al., 2021). Overall, these findings collectively suggest that the neurocomputational mechanisms involved in extracting numerosity are not limited exclusively to the parietal and frontal regions. Instead, evidence from computational modeling using CNNs suggests that regions along the ventral stream, including the IT cortex, may also play a role in representing numerical information.

In my study, I successfully identified regions in both the dorsal and ventral streams of the brain that are involved in representing numerosity. This finding suggests that numerosity processing is not confined to a single location but distributed across multiple cortical regions. Similar to how sensory maps operate (Young, 1998), it is plausible that different cortical locations in the brain function in parallel to analyze numerical information. Given the distinct functional roles observed in various cortical regions, such as object recognition, attentional control, decision making, and mathematical cognition, it is plausible that quantity processing plays a guiding role in these different cognitive functions. This notion is supported by previous research by Harvey and Dumoulin (2017a), who proposed that numerosity representation in each cortical location may serve different functional purposes. However, an alternative possibility is that the representation of numbers in ventral regions is a result of projections from the dorsal regions and that it develops as the effect of formal math education (Dehaene-Lambertz et al., 2018; Kersey et al., 2019). This connectivity-based framework suggests that the ventral regions receive input from the dorsal regions and uses this information. In a recent study by Conrad et al. (2022), compelling evidence was presented regarding the structural and functional connectivity of the NFA in the occipitotemporal cortex (OTC) with the parietal regions. For further investigation, it would be intriguing to explore whether the development of the NFA in its specific locations is influenced by differences in the connectivity of that region with the rest of the brain. This intriguing line of inquiry is reminiscent of a study conducted by Saygin et al. (2016), in which longitudinal scanning of children before and after they acquired reading skills revealed that early connectivity patterns play a crucial role in shaping the functional development of the visual word form area (VWFA). Drawing parallels from this study, exploring the potential influence of pre-existing connectivity on the developmental trajectory of the NFA could provide valuable insights into the mechanisms underlying numerical processing in

the ventral stream.

**Representations of non-numeric features are mostly restricted to the early visual area**

The results obtained from the searchlight analysis revealed an interesting pattern of brain activation. Specifically, the representation of non-numeric features was predominantly localized in the early visual area, while the representation of numbers spread across the cortex. These findings were further supported by the ROI analysis, which corroborated the results obtained from the searchlight analysis. It is noteworthy that only density showed amplified representation in the IPS, which could be attributed, at least in part, to its correlation with number. In my experimental design, density exhibited the strongest correlation with number among other non-numeric features. The significant representation of density detected solely through the ROI analysis could indicate a non-focal representation of density in the IPS as searchlight analyses have a tendency to bias towards identifying focal representations (Bulthé et al., 2019). Therefore, the detection of density representation only through the ROI analysis may result from a distributed representation of density in the IPS. Moreover, the total field area remained significant in the IPS, ventral occipital (VO) region, and parahippocampal cortex (PHC), without showing amplification along the dorsal or ventral stream. In contrast, it was specifically numerosity that exhibited amplified representation along the dorsal and ventral stream.

Quite surprising compared to my initial expectations, I failed to find any evidence of average item area representation along the dorsal or ventral stream using ROI or searchlight analysis. Previous research has demonstrated that observers can perceive the average size of an array of circles, even during brief displays lasting less than one-tenth of a second (Ariely, 2001; Chong & Treisman, 2003). However, the perception of average size is a topic of debate, partly due to the uncertainty surrounding the existence of low-level size detectors, unlike the perception of average motion, position, or orientation (Whitney & Leib, 2018). In a study by Castaldi et al. (2019), no evidence of information regarding the average item area was found in the pattern of neural activity across the examined regions, even when participants directed their attention to the average item area. These findings suggest that the neural mechanisms underlying the representation of average item area may differ from those involved in analyzing the size of individual objects, which has been shown to partially overlap with numerosity maps in parietal regions (Harvey et al., 2015). The perception of average item area has been a topic of inquiry, and some researchers propose that it, along with density perception, relies on texture processing mechanisms rather than the identification of individual items (Im & Halberda, 2012). This suggests that instead of processing each item individually, the brain processes the arrangement and patterns of elements in a visual scene to extract information about

average size or density. Recent findings propose that the ventral visual pathway may primarily represent local features rather than shapes (Ayzenberg & Behrmann, 2022). In a series of studies, researchers transformed naturalistic object images into texture-like representations called "texforms." These texforms preserved the visual statistics of the images while distorting the shapes of the objects, rendering them unrecognizable to human observers. Despite the lack of object recognition, these unrecognizable texform images evoked similar large-scale organizational patterns as real objects in the ventral pathway. The neural topography of the ventral pathway corresponded functionally to dimensions of animacy and real-world size (Long et al., 2018; R. Wang et al., 2022). Another study by Cant and Xu (2012) revealed that object ensembles and surface textures produce equal release from adaptation, suggesting similar processing mechanisms in the parahippocampal place area (PPA). These findings further support the notion that ensembles and textures are processed in a comparable manner. Based on these results, it might be expected that the representation of average item area would be present in the ventral pathway. However, in my study, I did not identify any specific region in the ventral stream that represented the average item area. The absence of average item area information could be attributed to a lack of attention. In a study conducted by Jackson-Nielsen et al. (2017), they used a modified inattentional blindness paradigm and observed that color and size ensembles were not perceived when attention was not directed towards them. This may explain why in my study I could find average item area representation in the ventral pathway. However, further investigation is needed to find out whether in presence of attention to this feature, I could find evidence of information regarding the average item area in the ventral pathway or not.

**The geometry of representation is similar across the dorsal and ventral regions (but not in retinotopically organized regions)**

Through the use of MDS on the dissimilarity matrices obtained from various regions of interest, I was able to unveil a condensed, low-dimensional representation of the underlying similarity structure present within these matrices. I visualize the first two dimensions of the MDS output. Closer proximity of the points to each other within this visualization correspond to patterns that exhibit greater similarity, while those situated farther apart indicate patterns with comparatively lesser resemblance (Nili et al., 2014). An evident pattern in the visualization resulting from MDS is a clear separation of stimuli with larger total field area and smaller total field area in the early visual areas (V1, V2, and V3). While this clear separation is evident in the ventral regions up to mid occipito-temporal regions (VO1 and VO2), this clear separation is replaced by a curved structure in the parietal region (IPS 12, IPS 345). In a recent study, conducted by (Nelli et al., 2023), a similar curved structure was identified around a midpoint from the MDS on BOLD data originating from the posterior parietal cortex (PPC). This analysis was conducted while subjects were

engaged in a comparison task involving rank-ordering novel objects. They argued that PPC encodes items at the extremes of the rank-ordered list, where comparisons are unambiguous, and middle items, where comparisons become less clear, with distinct neural codes for each. A similar mechanism might be at play in my case. It's plausible that subjects encoded the extreme numbers (9 and 29) differently from the intermediary numbers in the list (10 and 17) within the parietal regions.

The presence of this curved manifold enveloping a midpoint, as extracted from MDS on RDMs derived from the IPS, might reveal dimensions of numerosity perception that hold cognitive relevance. The first dimension arising from the MDS process can be seen as an absolute mapping of numbers onto space (Hubbard et al., 2005). The functional relevance of the second dimension remains unknown. According to Summerfield et al. (2020), the posterior parietal cortex is capable of encoding abstract relational information among stimuli, whether situated in physical space (Behrens et al., 2018) or within an abstract conceptual space (Freedman & Assad, 2006), all within a low-dimensional framework. This efficient, low-dimensional encoding proves particularly beneficial for adapting to novel situations. Thus, the presence of two distinct neural pathways, as revealed by our MDS analysis of ROIs, can likely be explained by the dimensionality of neural data. The ventral stream is responsible for creating high-dimensional representations crucial for identifying objects, whereas the dorsal stream focuses on forming low-dimensional representations that are optimal for abstraction and generalization. In the realm of numbers and magnitudes, it has been proposed that there could be overlapping neural codes for space, time, and number (Sheahan et al., 2021; Walsh, 2003), which assist in grasping the analogical connections between abstract concepts. Notably, this representation is one-dimensional, with numbers arranged along a left-to-right oriented 'mental number line,' where smaller numbers are positioned on the right and larger numbers on the left (Hubbard et al., 2005). Summerfield et al. (2020) further suggest that the acquisition of structural knowledge is intrinsically linked to action, similar to the function of neurons in the Lateral Intraparietal area (LIP), which serve as information accumulators to direct behavior (Mazurek et al., 2003; Roitman & Shadlen, 2002). Similarly, during a task where participants were asked to generate random numbers, they tended to move their eyes to the left when thinking of larger numbers and to the right when considering smaller numbers (Loetscher et al., 2010). Consequently, the observed curvature in the MDS may reflect a decision variable essential for behavior guidance; in terms of numbers, this could relate to the degree of being towards the left or right mental number line (or being smaller or larger). Notably, a recent CNN model, specially trained for a numerosity classification task, displayed a similar curved structure in its final layer (Mistry et al., 2023). This alignment suggests that a curved manifold could be functionally relevant in the context of comparative decision-making. These representations seem notably absent in the ventral retinotopic regions of the brain.

However, extending my analyses to non-retinotopic regions representing numerosity, as identified through the searchlight analysis, I found that the same curved manifold extended not only to lateral parietal cortex (in regions NPC1 and NOP2), but also, notably, to the lateral occipito-temporal (NTO) areas. This structural similarity could indicate the presence of important connectivity between the dorsal and ventral pathways (Conrad et al., 2022).

## 2.5 Conclusion

In summary, my study illustrates that visual numerosity is represented beyond the non-numerical visual features in a large set of regions. This representation is present already in early visual areas and spreads along both the dorsal pathway (up to the anterior parietal region) but also the ventral stream (reaching the lateral and inferior occipito-temporal cortex). This representation surpasses the boundaries of retinotopically organized regions and extends to higher-level regions in both streams. I also observed that the neural representational geometry was consistent across both the dorsal and ventral association areas, diverging significantly from that characterizing early numerosity representations in visual areas. The shared curved manifold in both streams, where numerical information is encoded alongside decision variables, underscores the potential importance of both dorsal and ventral associative regions in numerosity decision-making processes.

# CHAPTER 3

## Investigating spatio-temporal representation of visual numerosity of the human brain

# Chapter 3

# Investigating spatio-temporal representation of visual numerosity of the human brain

## 3.1 Introduction

Over the past two decades, the scientific community has accumulated substantial evidence of visual numerosity representation at the population level using fMRI (see the second chapter, for a review see also: Piazza & Eger, 2016). This process is frequently explored using visually presented dot arrays and has been investigated in various brain areas, including both dorsal, especially parietal regions (Castaldi et al., 2019; Eger et al., 2009; Harvey et al., 2013; Piazza et al., 2004), and also more recently ventral streams (Harvey & Dumoulin, 2017a, see also second chapter). While these studies illustrate where in the brain number representation may occur, the low temporal resolution of fMRI fails to provide insights into when numerosity information emerges in the brain. While previous finding on how the pattern of BOLD signal activity from fMRI across various brain regions can be simultaneously captured by both numeric and non-numeric features of the visual set of dots (Castaldi et al., 2019, see also second chapter) are quite convincing that number uniquely contributes to the variance of BOLD signals across different regions of the brain, neuroimaging results remain inconclusive regarding whether number is encoded directly (thus can be considered as a "primary" visual feature) or not. It is well-known that the visual system involves both a feedforward flow of information from early visual areas to higher brain regions and feedback from higher brain regions to the early visual areas (Lamme & Roelfsema, 2000). Therefore it is possible that even if number appears as being represented independently from other features in a certain region, e.g., the intraparietal sulcus (IPS), that representation could be the outcome of a combination of other non-numeric features encoded earlier on in another region, e.g., early visual areas. At the same time, observing a separate representation of number and other visual features in early visual areas does not guarantee that the pattern of response represents what is computed early in time and that is not due to a back-projection

from higher-level brain areas. Hence, understanding the timing of the representation of numeric and non-numeric features becomes crucial to be more specific about how number is encoded in the brain. If numerosity information is encoded before other non-numeric features from which number can be computed, then this is strong evidence that number could be considered a primary visual feature. On the contrary, if numerosity information is encoded after these features, it is possible (but not necessary) that it is derived by the combination of non-numeric information, indicating an indirect encoding process. To the best of my knowledge, currently only few studies directly investigated the temporal representation of visual number in human adults using electroencephalogram (EEG; Fornaciai et al., 2017; Gebuis & Reynvoet, 2013; Park et al., 2015). Contrary to the study by Gebuis and Reynvoet (2013), which found no evidence of numerosity-related effects in the ERP signals of EEG, both Park et al. (2015) and Fornaciai et al. (2017) provide evidence that a significant portion of variance in the ERP signal can be attributed to numerosity. The absence of numerosity information in the ERP signal in Gebuis and Reynvoet (2013) study may be attributed to problematic stimuli design, where there was much larger variation in non-numerical features than in numerosity. This difference could result in masking the effect of numerosity, as it has less variance and is therefore less detectable. On the other hand, although Park et al. (2015) and Fornaciai et al. (2017) provided evidence for the representation of numerosity very early in time, their results remain inconclusive, as they only explored the timing of numerosity and two mathematically constructed variables, "size" and "spacing" (see the second chapter of the thesis). This raises the question of how other perceptually relevant non-numeric features of the visual set of dots are represented in the brain and whether these features can account for the numerosity information observed or not.

In this study, similar to previous chapter, I employed time-resolved representational similarity analysis (Kriegeskorte et al., 2008) to model the unique contributions of numeric and all perceptually relevant non-numeric features to the magnetoencephalography (MEG) data. The rationale behind this approach is that if numbers are encoded before non-numeric features that could indirectly contribute to the representation of numbers, then numbers should be considered primary visual features.

While time-resolved results from EEG or MEG can provide a clear picture of the temporal representation of numeric and non-numeric features of visual stimuli, the cortical source of each feature remains a question due to the limited spatial resolution of EEG or MEG. While some studies (Fornaciai et al., 2017; Park et al., 2015) have attempt to support the claim that the early visual area acts as the source for extracting numerosity information from visual stimuli by relying on the observation that a significant proportion of the variance in the ERP can be attributed to numerosity on the medial occipital channel of EEG, linking scalp topographies to underlying source locations proves challenging due to the well-established phenomenon that electrical potentials from various sources blend at the level of scalp recordings (Baillet, 2017).

Other studies that employed steady-state visual evoked potentials (SSVEP), a brain response to stimuli flickering at a specific frequency, in children (Park, 2018) and adults (Guillaumé et al., 2018; Lucero et al., 2020; Van Rinsveld et al., 2020) have made a similar claim regarding the underlying source of numerosity. However, their results are again unfounded as they are solely based on the scalp topography of the EEG response. Hence, the temporal unfolding of numeric and non-numeric features across different brain regions remains a subject of debate. While fMRI data (as discussed in the second chapter) can give us insights into where numerosity is represented in the brain, and MEG data can provide a clear temporal representation of numeric and non-numeric features across the brain, we lack a clear understanding of the temporal unfolding of each feature within the regions that fMRI results suggest numerosity is encoded.

Here and in the rest of this chapter, I explored the temporal dynamics of various retinotopic regions (L. Wang et al., 2014) along both the dorsal and ventral streams using MEG-fMRI fusion based on representational similarity analysis (Cichy & Oliva, 2020). Employing a model-based MEG-fMRI fusion allowed to determine where in the brain and to what extent each numeric and non-numeric feature is encoded across the time course of trials. This approach enabled us to ascertain whether numbers are encoded before non-numeric features within the predefined brain regions. Moreover, it provided insight into the latency differences across the brain regions.

## 3.2 Methods

### 3.2.1 Participants

Thirty-nine healthy adults (twenty-one females, average age 25.1 years) with either normal vision or vision corrected participated. The sample size for our study was established based on a prior study by Castaldi et al. (2019), which involved twenty healthy adults. Our chosen sample size is comparable to or even exceeds the typical sample sizes employed in similar experiments in the field (e.g. Fornaciai et al., 2017; Gebuis & Reynvoet, 2013). The University of Trento's ethics committee in Italy approved the study, and all participants provided written informed consent and were compensated for their participation. To keep head movement below 10 mm, data from six participants were omitted from the final analysis. Additionally, three more participants were excluded because of poor behavioral performance (accuracy below 60%). This resulted in a final sample of thirty subjects (17 females, average age 25.6 years).

### 3.2.2 Stimuli and procedure

Before the MEG experiment, participants were introduced to the task through 20 practice trials conducted outside the MEG environment. During the MEG scanning

session, participants were presented with arrays of black dots on a mid-gray background. These dots orthogonally varied in terms of number, average item area, and total field area. There were 32 conditions resulting from the combination of four numerosities (six, ten, seventeen, or twenty-nine dots), four average item areas (0.04, 0.07, 0.12, 0.2 visual square degrees), and two total field areas (defined by a virtual circle with diameters of approximately 9 or 13.5 visual degrees; Figure 3.1). The selection of these numbers and average item areas was based on their perceptual discriminability, as established in previous research (Castaldi et al., 2018, 2019). The total field areas were chosen to ensure suitably sparse arrays of dots (1 dot/vd$^2$) to be within the numerosity estimation regime and not within the density estimation regime (Anobile et al., 2013).

In each trial, a set of dots was presented for 500 ms and displayed over a wide thin red fixation cross (the sample set). Participants were asked to estimate the number of dots and hold that information in memory until the appearance of the next set, after a variable interstimulus interval (ISI) of 500 or 800 milliseconds (Figure 3.1). When the color of the fixation cross changed from red to green, participants were instructed to compare the number of dots in the current set (the match set) with the previous set and indicate whether it was larger or smaller by pressing one of two buttons, following provided instructions. After a delay of 1.8 seconds, the next trial began. Match sets differed by approximately 2 just noticeable differences (JNDs) in numerosity from the previous sample stimulus, based on the average numerosity JND estimated from a separate group of healthy adults (Castaldi et al., 2018). The other features (total field area and average item area) remained constant. Match trials occurred approximately 20% of the time.

The experiment consisted of twelve runs, each containing four blocks. Each block included thirty-six trials, comprising four match trials and thirty-two sample trials, one for each of the thirty-two conditions (4 number × 4 average item area × 2 total field area). After the sixth run participants' response assignments were switched. Brief practice sessions were conducted at the beginning of the experiment and after the hand assignment change. Each run had a duration of approximately 3.5 minutes.

### 3.2.3 MEG recordings and preprocessing

Subjects' brain activity was recorded at the Center for Mind/Brain Sciences at the University of Trento using a MEG system comprising 306 channels (204 planar gradiometers and 102 magnetometers) manufactured by Elekta-Neuromag Ltd. in Helsinki, Finland. The data was acquired at a sampling rate of 1000 Hz and underwent online band-pass filtering within the frequency range of 0.01–330 Hz. The participants were seated upright within a room that was shielded against magnetic interference (AK3B, Vakuumschmelze, Hanau, Germany). Prior to the recording session, the unique head shape of each participant was measured using a Polhemus FASTRAK

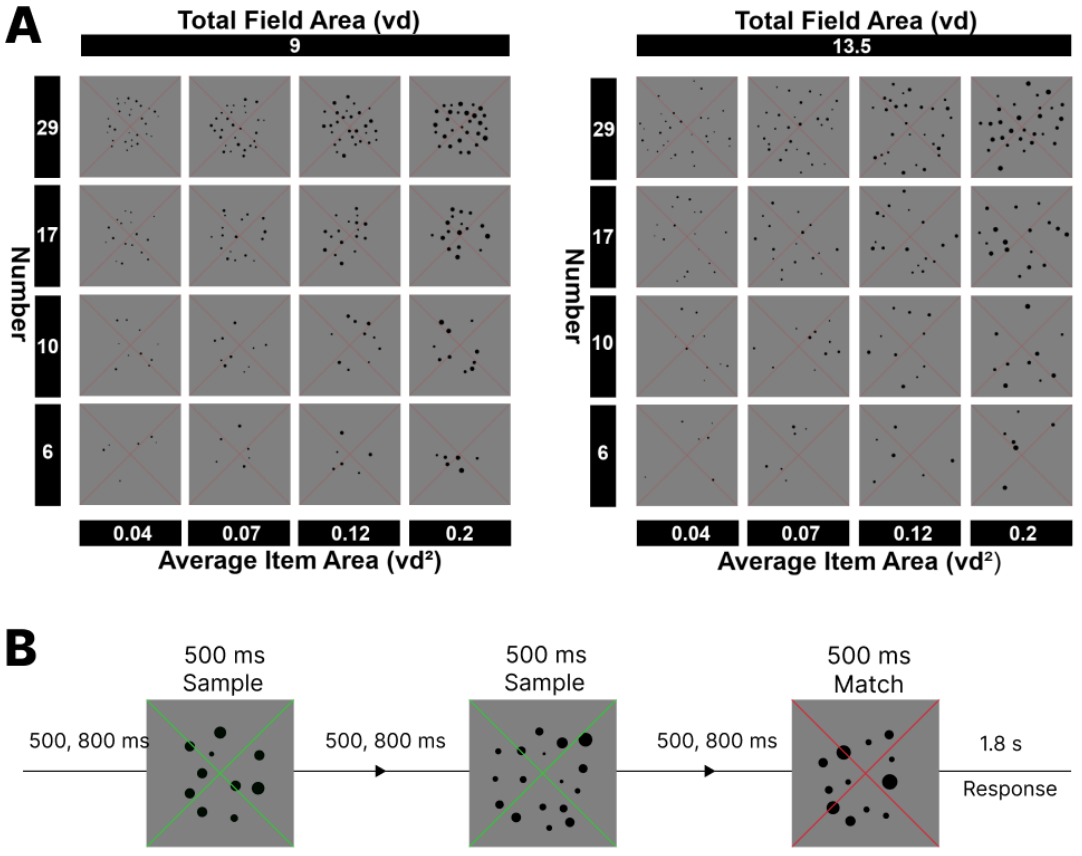



FIGURE 3.1: (A) A visual representation of the entire collection of stimulus conditions. The dataset encompasses orthogonal variations in multiple aspects, including the number of items (6, 10, 17, 29), the average item area (0.04, 0.07, 0.12, and 0.2 visual square degree), and the total field area defined by imaginary circles with diameters of 9° and 13.5°. (B) Visual representation of the chronological sequence of trials during the scanning process. Participants were instructed to focus on the number of dots and retain this information until the subsequent dot set was displayed. Periodically, the fixation cross's color transitioned from red to green. When this color shift occurred, participants were instructed to assess whether the number of dots in the current set, known as the match stimulus, was greater or lesser than the previous one, and indicate their choice by pressing a button with their left or right hand.

3D digitizer (Polhemus, Vermont, USA). This process involved acquiring three fiducial points (nasion and pre-auricular points), five head position indicator (HPI) coils (one on each of the left and right mastoids and three on the forehead), and approximately 100 more locations distributed across the subjects' skull. At the start of each run, head positioning inside the MEG helmet was measured by inducing a non-invasive current through the HPI coils. Additionally, both horizontal and vertical electro-oculograms (EOGs) were recorded, along with an electrocardiogram (ECG), with the intention of later offline eliminating artifacts related to eye movements and heart activity. For stimulus presentation and control, Psychtoolbox 3 (Brainard, 1997) was employed. Each visual stimulus image was rear-projected onto a screen positioned at a distance of 120 cm from the participant's eyes using a VPixx PROPixx projector.

The offline raw MEG data underwent a visual inspection to identify and eliminate noisy channels. Subsequently, I employed the MaxMove function of the Elekta Maxfilter software to remove head motion and to denoise the data using Maxfilter Signal Space Separation (Taulu et al., 2003). Following the maxfiltering, I proceeded with additional preprocessing using MNE-Python version 0.24.1 (Gramfort, 2013). First, the data underwent band-pass filtering within the range of 0.05 to 330 Hz and band-stop filtering at 50 Hz and its harmonics for line noise removal. To eliminate artifacts caused by eye movement and heart rate, independent components of the MEG data were computed and correlated with the EOG and ECG signals. Components displaying significant correlations were subsequently removed through manual inspection. Subsequently, to enhance signal-to-noise ratio (SNR) and lower computational demands, I applied temporal smoothing on signals using Savitzky–Golay filter (Acharya et al., 2016, with order 4 and frame size 25, ) and downsample the data to 200 Hz (Grootswagers et al., 2017). Following these preprocessing steps, the data was epoched into 0.9-second trials, starting 100 ms before the stimulus onset. Each epoch was then normalized by subtracting the baseline period mean. Given that magnetometer and gradiometer sensors have different measurement units, I opted to exclusively use gradiometer data for my subsequent analysis. This choice aligns with prior studies, such as Ritchie et al. (2015), which exclusively focused on gradiometer measurements, as well as studies like Kaiser et al. (2016) that concentrated solely on magnetometer data. Additionally, some studies, such as Proklova et al. (2019), have chosen to report results separately for each sensor type.

### 3.2.4 Time Resolved Multivariate Representational Similarity Analysis (RSA)

To explore whether and when the representations of numerical and non-numerical features of the stimuli can be disentangled from the brain signal I employed representational similarity analysis combined with semipartial correlation (Kriegeskorte & Kievit, 2013; Kriegeskorte et al., 2008).

To create neural representational dissimilarity matrices (RDMs), I organized the preprocessed MEG data into pattern vectors, each encompassing MEG data from all selected channels for each trial and time point. Following this, I averaged data from trials of the same conditions into single trial representations. Consequently, this process yielded 32 distinct MEG patterns, each corresponding to one of the 32 conditions, for each time point (100 ms before to 600 ms after stimulus onset). To quantify the similarity between all pairs of these 32 patterns, I employed Pearson correlation. I then subtracted these correlation values from 1 to generate 32 x 32 MEG RDMs for each time point. Following the approach outlined in the previous chapter, I then conducted a semipartial correlation analysis (using Pearson correlation) to assess the extent to which the dissimilarity structure in MEG patterns could be attributed to multiple predictor or model matrices that capture key quantitative dimensions of

the stimuli. These dimensions include number, average item area, total field area, total surface area, and density. These five model RDMs quantitatively represent the logarithmic distance between pairs of stimuli in terms of number, average item area, total field area, total surface area, and density. The semipartial correlation between a vectorized neural RDM and the specific model RDM under examination measures the unique variance shared between the neural RDM and the chosen model RDM while partialling out the effect of all other models RDM from the neural RDM. A diagram outlining this procedure is presented in Figure 3.2 The analysis was implemented using the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016) and custom-written code in MATLAB R2019 (The MathWorks, Inc., Natick, MA).

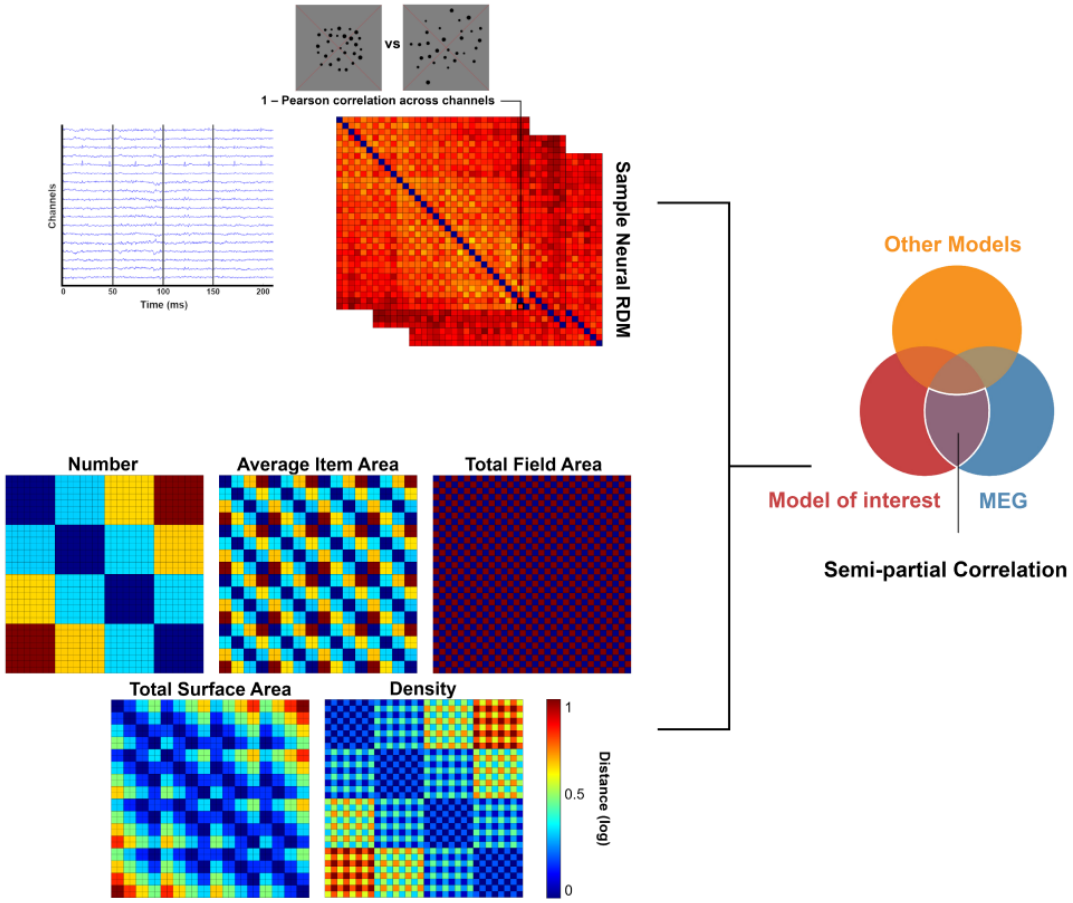



FIGURE 3.2: Representational dissimilarity matrices (RDMs) extracted from MEG signals were subjected to a semipartial correlation analysis. MEG RDMs were generated by computing the 1 – Pearson correlation between channel activations across time points for every pair of images. In the semipartial correlation analysis, five model RDMs were used as predictors, representing the logarithmic distance between pairs of stimuli based on number, average item area, total field area, total surface area, and density.

### 3.2.5 Sensor-space Searchlight Multivariate RSA

In order to identify sensors over time where the independent contribution of five model RDMs to the neural RDM could be observed, I conducted a sensor-space

searchlight analysis (Kriegeskorte et al., 2006; Proklova et al., 2019). Applying the method described in the preceding section, I employed semipartial correlation analysis (Pearson correlation) to evaluate the degree to which the dissimilarity structure in MEG patterns be explained by the model matrices. To achieve this, I designated a neighborhood for each MEG channel comprising its 20 nearest MEG channels. Subsequently, I conducted time-resolved RSA for each MEG channel, with the analysis being restricted to the data from its neighboring channels. I proceeded to average the results in 150 ms bins, producing a single map representing the grand average correlation coefficient for each feature.

The analysis was carried out using the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016) alongside custom-written code in MATLAB R2019 (The MathWorks, Inc., Natick, MA). The results were then visualized using the *ft_topoplotER* function in the FieldTrip Toolbox (Oostenveld et al., 2011).

### 3.2.6 Time-Frequency Resolved Multivariate RSA

I employed time-frequency decomposition to investigate whether the representation of numbers occurs in a distinct frequency band compared to other non-numeric features. These findings could be crucial to indicate whether number is encoded separately from other non-numeric features. Following Xie et al. (2022), I produced time-frequency decompositions of the preprocessed MEG time series using Morlet wavelets for each trial and sensor. The wavelets used in the analysis had a consistent length of 2,600 ms and were logarithmically distributed across 40 frequency bins ranging from 2 Hz to 30 Hz. Absolute power values for every time point and frequency bin were derived by taking the square root of the resulting time-frequency coefficients. These power values were then normalized to represent relative changes, expressed in decibels (dB), concerning the pre-stimulus baseline (–100 ms to 0 ms in relation to stimulus onset). This analysis was carried out using custom-written MATLAB R2019 code (The MathWorks, Inc., Natick, MA). I decompose the MEG time series using a custom function developed by Xie et al. (2022) (available at https://github.com/siyingxie/VCR_infant/blob/main/code/timefrexdecomp.m).

### 3.2.7 Temporal Generalization Analysis

I employed temporal generalization analysis (King & Dehaene, 2014) to better explore the potential evolution of the neural representation of numbers over time. This analysis allows making inferences on the level of stability of the neural code in time. In particular, when it is possible to decode a given information across different time points this indirectly proves that the neural representational format of that information remains sufficiently constant in time. On the contrary, when decoding cannot be extended in time, this indicates that the neural representational format undergoes important transformations. To conduct this analysis, I first randomly combined every four trials into pseudo-trials in order to enhance the overall signal-to-noise ratio

(Guggenmos et al., 2018). Subsequently, I trained a support vector machine (SVM) classifier to decode the four numbers (6, 10, 17, and 29) at a specific time point and then tested the same classifier across all other time points separately. The preceding steps were repeated for a total of twenty randomized assignments of trials to pseudo-trials (permutations). Finally, I computed the average decoding accuracy across all permutations. The analysis was carried out using the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016) and LIBSVM (Chang & Lin, 2011).

### 3.2.8 Model-based MEG-fMRI Fusion

To unfold the temporal dynamics of different brain regions involved in processing numerosity and the non-numeric features I used model-based MEG-fMRI fusion (Cichy & Oliva, 2020). The RSA-based MEG-fMRI fusion approach enables us to ascertain whether and when (in time) the representational structure in a specific brain region, as determined by fMRI, aligns with the representational structure derived from time-resolved MEG signals. When there is a correspondence between the fMRI RDM of a brain region and the MEG RDM at a particular time point, it implies a shared representational format within that brain area at that specific moment in time (Cichy, Pantazis, & Oliva, 2016; Cichy et al., 2014). Here, building upon the work of Hebart et al. (2018), I expanded this approach to incorporate different models in the fusion process. This extension allowed us to spatiotemporally resolve the processing of both numeric and non-numeric information.

For the model-based MEG-fMRI fusion, three distinct types of RDMs were employed: (I) five model RDMs representing number, average item area, total field area, total surface area, and density, as previously described, (II) MEG RDMs from individual subjects, as detailed earlier, and (III) five group-averaged RDMs derived from thirty-one participants in five regions of interest (ROIs) as outlined in the second chapter. In my fMRI study, I employed the exact same task as in the current study, except for the timing of the stimuli, which were presented at a faster pace in the MEG experiment. I used t-statistics from the first-level analysis to derive the neural RDM. This was achieved by calculating the correlation distance between activation patterns for every pair of conditions, done separately for each region of interest (ROIs: V13, V3ABV7, IPS15, VO, and PHC).

Twenty-three participants from the fMRI study overlapped with the participants in this study. Consequently, I used the group-average RDMs from the fMRI study as the most reliable estimate of the actual pattern dissimilarity (Nili et al., 2014). A schematic of this process is shown in Figure 3.3. For subsequent analysis, I extracted the lower triangular component of each RDM, excluding the diagonal, and vectorized them. These vectors were referred to as representational dissimilarity vectors (RDV). To identify the shared variance among three desired RDVs (MEG RDV, fMRI RDV, and number RDV) while partialling out the effect of other RDVs (RDVs related to non-numeric features, including average item area, total field area, total surface

area, and density), I initially regress out the effects of all five model RDVs. This process allowed us to isolate the unexplained portion (residual) within the MEG RDV. Subsequently, I calculated the semipartial correlation coefficient between the fMRI RDV and the MEG RDV, while partialling out the effects of all non-numeric RDVs and the residual RDV from the previous step. The resulting semipartial coefficient illustrates the shared variance among MEG, fMRI, and number RDV, as depicted in the Venn diagram in Figure 3.3.

### 3.2.9 Multidimensional Scaling (MDS)

While the results from RSA reveal how each feature contributes to the variance of my data in a hypothesis-driven manner, I also explore the latent structure of my data using a data-driven approach with MDS (Kruskal, 1964). This approach was applied directly to the RDM, and subsequently I visualized the first two dimensions of the MDS output. With MDS I visualize the organization of stimuli on a two-dimensional plot, where the distances between them accurately reflect the differences in the neural response patterns they evoked. Stimuli positioned closely together in these representations represent similar neural response patterns (Nili et al., 2014). The MDS analysis was conducted using the MATLAB function *cmdscale* on the group-average RDM across participants spanning various time points from the onset of stimulus presentation to 600 ms. The group-average RDM is calculated by averaging the RDMs from all thirty subjects. This group-level RDM is more accurate and less susceptible to noise compared to the RDMs obtained from individual subjects (Nili et al., 2014).

### 3.2.10 Statistical Testing

Throughout this chapter, statistical significance was evaluated through a one-sample t-test against zero across subjects. To control for multiple comparisons, the results were corrected using threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009), employing Monte Carlo simulations with 10,000 permutations, as implemented in the CoSMoMVPA MATLAB toolbox (Oosterhof et al., 2016). The resulting statistics were then thresholded at $p < 0.01$ (one-tailed) to determine significance. For the searchlight analysis, significance was evaluated independently for each time point using the TFCE method. This method helped identify sensors where the contribution of a specific predictor to neural dissimilarity was significantly above zero.

## 3.3 Results

Thirty healthy adult volunteers were presented with arrays of dots varying in the number of items (6, 10, 17, 29), average item areas (0.04, 0.07, 0.12, 0.2 visual square degrees), and total field areas (9 or 13.5 visual degree diameter) while undergoing

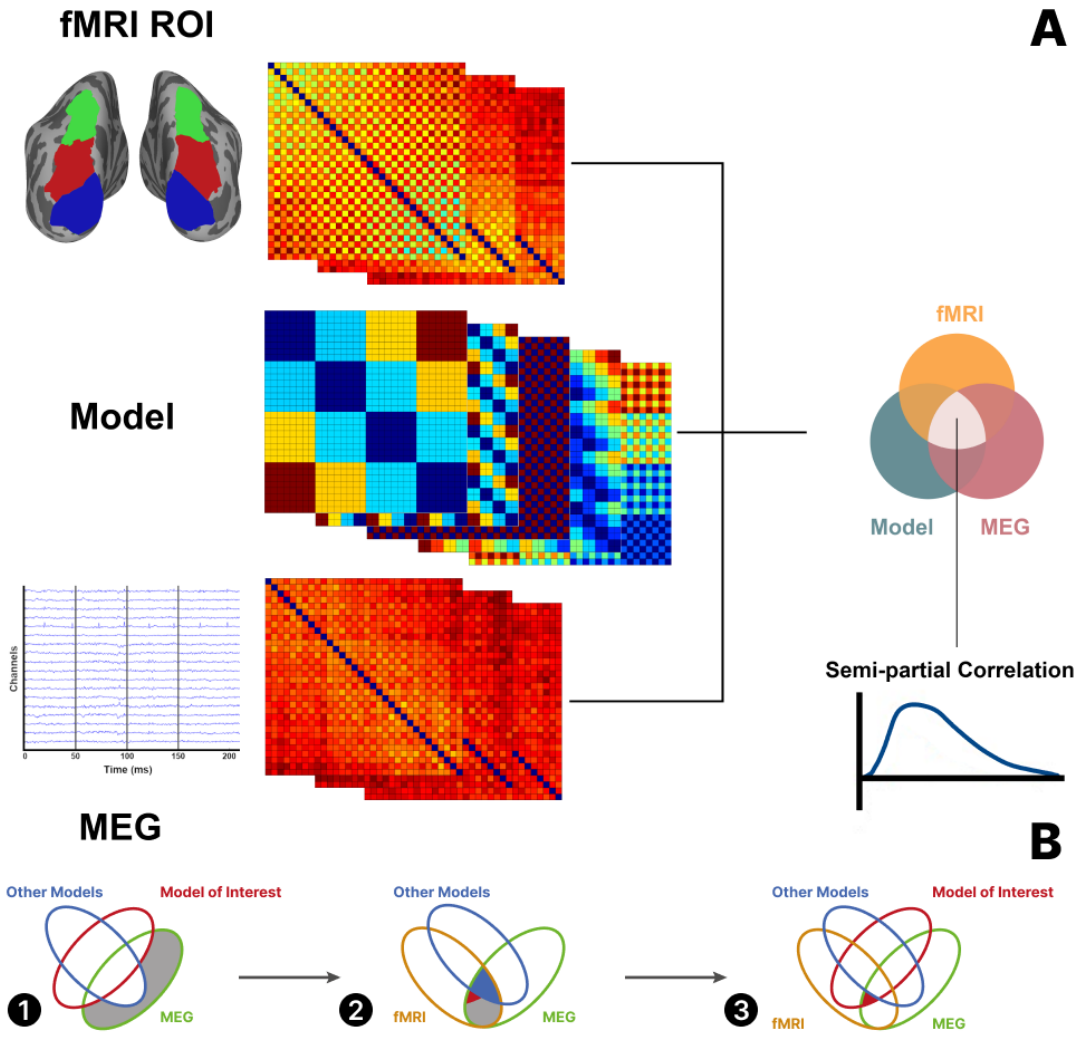



FIGURE 3.3: (A) The current model-based MEG-fMRI fusion formulation involves capturing the shared variance among three dissimilarity matrices: (1) a group-averaged fMRI RDM created from voxel patterns within a specific ROI, (2) a model RDM representing the expected dissimilarity structure related to a variable of interest (e.g., number), while excluding the influence of other variables of interest (e.g., average item area, total field area, total surface area, and density), and (3) an MEG RDM derived from MEG data at a specific time point. This analysis was conducted independently for each MEG time point, resulting in a time course of semipartial correlations for each ROI. (B) Identifying the shared variance between the fMRI, MEG, and the RDM representing the model of interest, while controlling for the impact of other model RDMs, involves a two-step process. First, the effects of other models and the models of interest are regressed out to obtain residuals. Second, the semipartial correlation between fMRI and MEG is computed, while also controlling for the effects of other models and the residuals from the initial step. The resulting semipartial correlation coefficient indicates the shared variance between fMRI, MEG, and the model of interest, with the influence of other models partialled out.

scanning in an MEG system. Participants were tasked with memorizing the number of dots from a sample stimulus and comparing it with a subsequently presented match stimulus, signaled by a change in the color of the fixation cross. Upon seeing the match stimulus, participants had to determine whether it was more or less numerous than the previous sample stimulus and indicate their response by pressing

a button with either their right or left hand. Participants demonstrated high overall performance (Mean=%77.81, SD=%5.83, Range=65.21% –88.54%), indicating their attentiveness and engagement in the task.

### 3.3.1 Time Resolved Multivariate Representational Similarity Analysis (RSA)

In order to characterize how numeric and non-numeric feature representations change over time, I employed time-resolved multivariate RSA. Figure 3.4 displays the results of the semipartial correlation between the neural and model RDMs from 100 milliseconds before the stimulus presentation to 100 milliseconds after it disappeared from the screen (-100 ms to 600 ms). Semipartial correlation ensures that the coefficients obtained reflect the variance uniquely explained by each model, while partialling out the influence of all other models. The findings reveal that the brain activity is significantly modulated by number over and above the other non-numerical features, with significant effects in three time windows: 115 – 140 ms (peak at 135 ms), 245 – 290 ms (peak at 260 ms), and 335 – 355 ms (peak at 340 ms) after stimulus presentation. The other non-numeric features, except for average item area, show significant effects either earlier (total field area: 65 – 600ms, peak at 120 ms, total surface area: 75 – 275 ms, peak at 165 ms) or later (density: 160 – 250, peak at 215 ms) than number. Thus, the representation of number preceded that of two specific pairs of non-numeric features, specifically total field area and density on one side, and total surface area and average item area, from the combination of which number can be indirectly inferred (see introduction).

### 3.3.2 Sensor-Space Searchlight RSA

Results from the sensor-space searchlight RSA (Figure 3.5) indicate that information related to number independent from other features is present in the MEG channels positioned over the occipital and parietal cortex already within 150 ms from stimulus presentation and is further amplified in time up to 450 ms. In contrast, information concerning other non-numeric features appears to be more pronounced in the MEG channels overlying the occipital cortex and instead of increasing in time they appear to be decreasing after 300 ms. In terms of differences in localization, these findings appear to align with those obtained from the searchlight analysis in my fMRI study (see second chapter), which demonstrated that while the representation of numbers is distributed across the cortex, non-numeric features are predominantly localized in early visual areas. In terms of timing the observation that pure number information peaks in the MEG channels beyond the strictly posterior ones within 150 ms suggests that numerical representations emerge very early and potentially also in regions that extend beyond primary visual cortex.

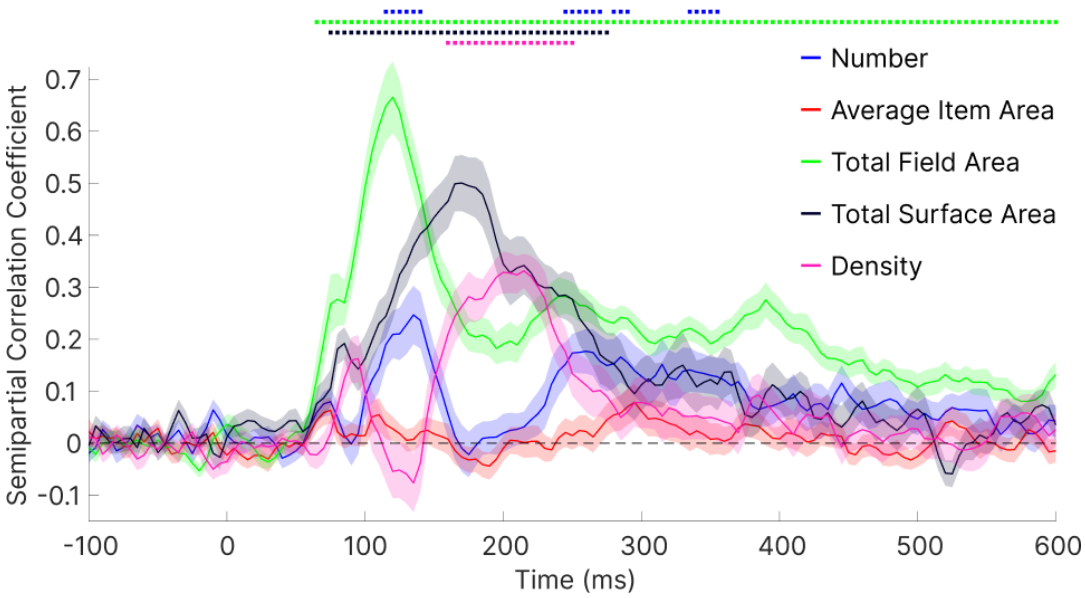



FIGURE 3.4: Semipartial correlation coefficients derived from the representational similarity analysis for number, average item area, total field area, total surface area, and density. Standard error of the mean (SEM) across participants is depicted as a shaded area. The horizontal coloured dots indicate significant time points with effects significantly exceeding zero (thresholded at $p < 0.01$, TFCE corrected).

### 3.3.3 Time-Frequency Resolved RSA

I then further expanded the analysis from the time domain to the time-frequency domain. The rationale behind this approach lies in the possibility that if numbers are represented in a distinct frequency range compared to other non-numeric features, it suggests that numeric and non-numeric features are represented through different functional networks. The time-frequency analysis, depicted in Figure 3.6, demonstrated that number is represented independently from other visual features in the lower beta range (12-17 Hz), whereas density and total surface area are represented in the higher beta range (17-30 Hz). Total field area is represented in a broader frequency range that encompasses both lower and higher beta. However, total field area varied orthogonally to number in my experiment. Thus the fact that it is represented in the same frequencies cannot be taken as an index for their functional dependency. Finally, the average item area did not exhibit significance within any frequency range.

### 3.3.4 Temporal Generalization Analysis

The results from the time-resolved RSA in the first section of this chapter demonstrates that number is represented independently from the other non-numeric features in three periods, and that in the first one (115 – 140 ms), peaking at 135 ms numerosity is unlikely computed based on the other visual features. However, in

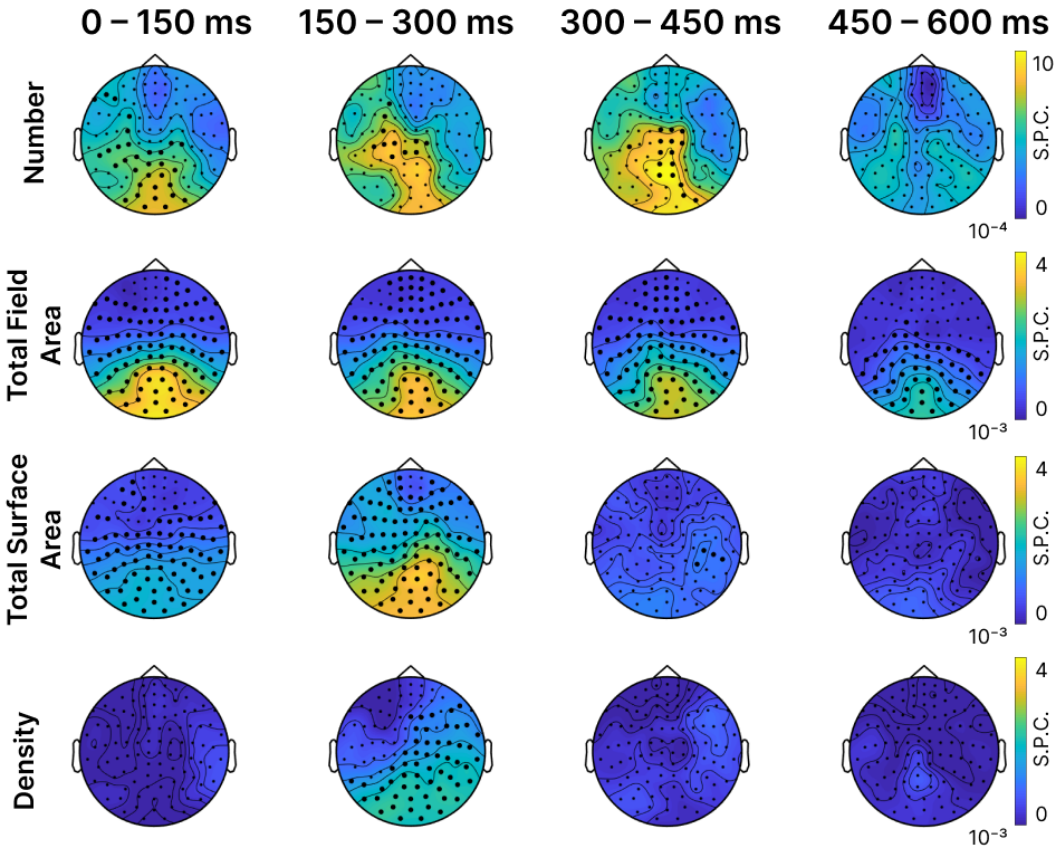



FIGURE 3.5: Semipartial correlation coefficients obtained from the representational similarity analysis demonstrate the contribution of numeric and non-numeric features on MEG patterns across channels (gradiometers) within four distinct time windows. Bold dots indicate sensors where semipartial correlation coefficients were significantly greater than zero (thresholded at $p < 0.01$, TFCE corrected). Here, I did not use the same scale for the number and the other three features as doing so would render the effect of number on sensors unrecognizable.

the second and third periods (245 – 290 ms and 335 – 355 ms), all non-numeric features, except average item area, are already encoded. This raises the possibility about whether numerosity undergoes important transformations in time.

To investigate the stability of number representation over time, I conducted a temporal generalization analysis (King & Dehaene, 2014). If the neural representational geometry of numerosity changes in time this would not allow cross-decoding in time. The results, depicted in Figure 3.7, revealed significant cross-temporal decoding especially after 100 ms (Figure 3.7), which is suggestive of a process that quickly stabilizes into a stationary representational format of numerosity.

### 3.3.5 Model-based MEG-fMRI Fusion

Prior fMRI studies have demonstrated that numerical information is represented across retinotopic regions in both the dorsal stream up to the Intraparietal Sulcus

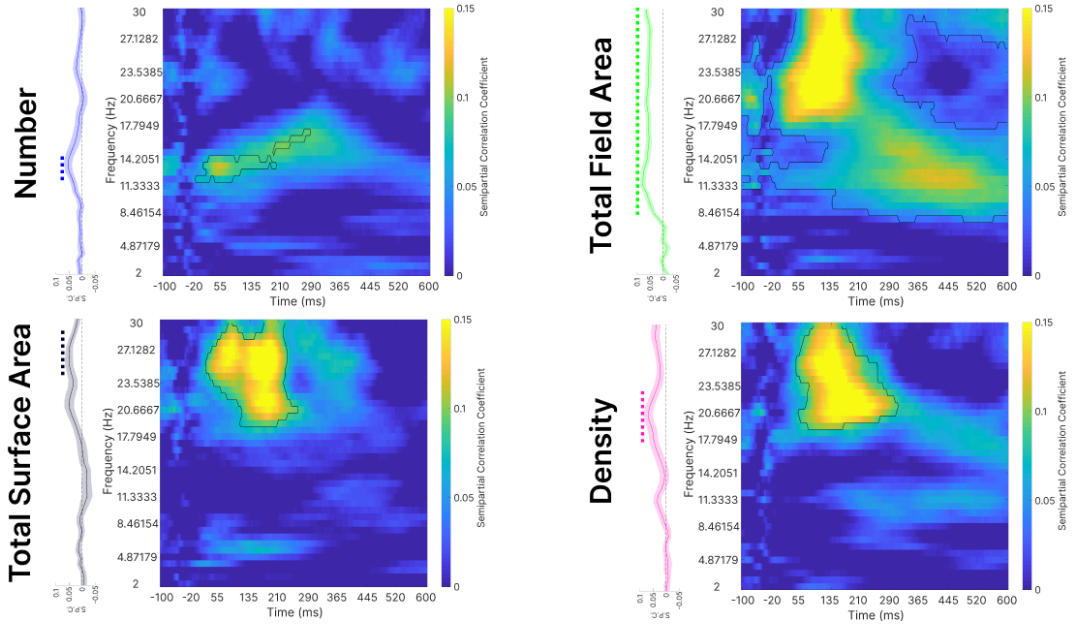



FIGURE 3.6: Semipartial correlation coefficient obtained from time-frequency resolved RSA for number, average item area, total surface area, and density. The outline reflects the significant clusters (thresholded at $p < 0.01$, TFCE corrected). The vertical figure next to each matrix represents the average of individual frequencies across time. The colored dots above them indicate significant frequencies (thresholded at $p < 0.01$, TFCE corrected).

(Castaldi et al., 2019, see also second chapter) and the ventral stream up to Ventral Occipital (see also second chapter). Results from the time-resolved RSA in the previous sections indicate that pure numerical information, independent of other non-numeric features, is represented very early in time. However, due to the limited spatial resolution of MEG and the sluggish nature of BOLD signals in fMRI, a detailed understanding of brain activity across time in individual regions of interest remains elusive. For example, the representation of numerosity observed in early visual cortex in the prior fMRI studies (Castaldi et al., 2019, see also second chapter) might either reflect a feedforward processing starting there but it might also be the result of feedback signals from higher-order cortical areas. To distinguish between these hypotheses, I employ a model-based MEG-fMRI fusion approach. The findings from the model-based MEG-fMRI fusion, illustrated in Figure 3.8 and Figure 3.9, showed that pure numerical information is equally rapidly represented across regions in the early visual area (V1-3), dorsal (V3ABV7, and IPS1-5) and ventral (VO, PHC) streams, preceding pairs of non-numeric features (total field area and density, or total surface area and average item area) which could indirectly contribute to the encoding of numerical information in all of them. The fact that I was not able to detect variations in the onset of numeric and non-numeric features across different regions, suggests the parallel emergence of these representations in these brain regions within the temporal precision afforded by my analysis approach. This interpretation is consistent with the findings of the sensor-space searchlight RSA, which tentatively

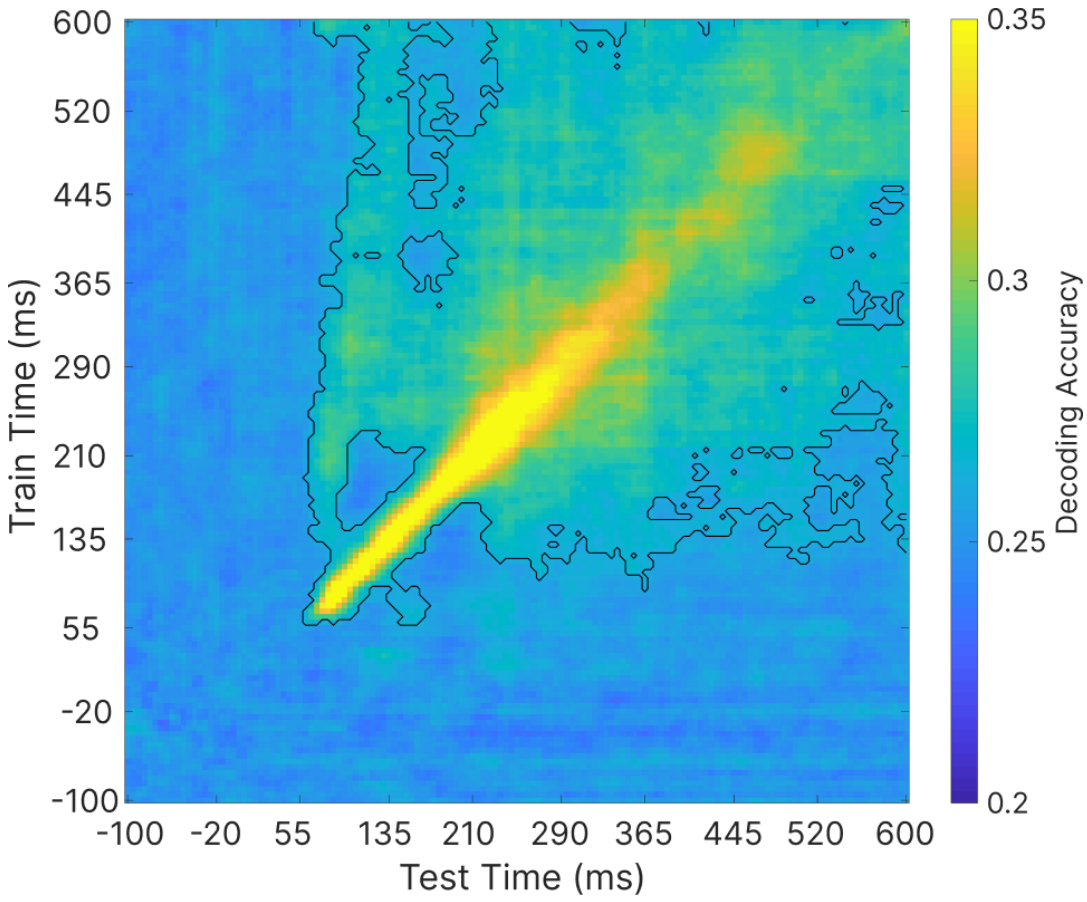



FIGURE 3.7: The temporal generalization matrix reflects the decoding accuracy (with a chance level of 0.25). The y-axis represents the training time during which the classifier was trained to classify among four numbers (6, 10, 17, 29), while the x-axis represents the test time when the classifier was tested to decode the numbers. The outline reflects the significant clusters (thresholded at $p < 0.01$, TFCE corrected).

suggest that the source of numerosity information spreads beyond strictly posterior sensors since very early in time. The results of fusion further showed that while numerosity information is present across all regions, it is particularly strongly present in IPS1-5 and VO, consistent with my fMRI study. Together, these results highlight a parallel and direct representation of numerical information along both the dorsal and ventral streams, with an increasing significance of numerosity information observed in IPS1-5 and VO.

### 3.3.6 Multidimensional Scaling

Finally, I employed multidimensional scaling (MDS) to derive a low-dimensional representation of the similarity structure. I applied MDS to the group-average RDMs derived from thirty participants calculated across various time points, ranging from the onset of stimulus presentation to 100 ms after it vanished from the screen, up to 600 ms (see also the attached Video 1).

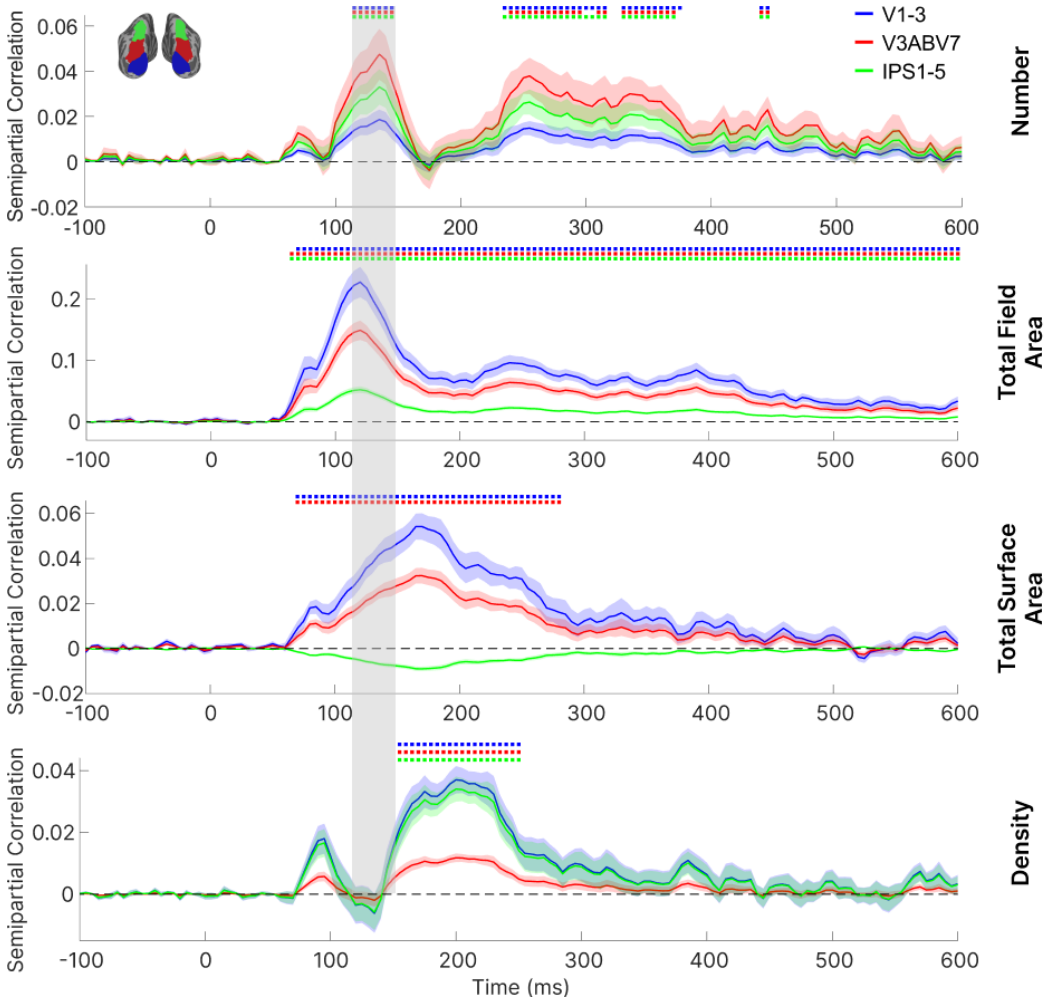



FIGURE 3.8: Time courses of semipartial correlation coefficient depicting the shared variance between model of interest (number, average item area, total field area, total surface area, density), fMRI RDM from the dorsal stream and MEG RDM. Standard error of the mean (SEM) across participants is depicted as a shaded area. The horizontal coloured dots indicate significant time points with effects significantly exceeding zero (thresholded at $p < 0.01$, TFCE corrected). The gray rectangle overlaid on each figure highlights that numbers are represented before pairs of non-numeric features (total field area and density or total surface area and average item area) that could indirectly contribute to number computation. It's important to note that the time courses of the average item area are not depicted here, as they did not show significance in any region.

The MDS results reveal a clear rank-ordering of numbers, starting around 100 ms post-stimuli. Furthermore, a curved pattern becomes apparent at specific time points during the second period when numerosity becomes significant (see Figure 3.10 for a snapshot of that time point, 275 ms, 290 ms, and 300 ms). This curved structure bears a resemblance to the patterns observed in the parietal region in my earlier fMRI study (see the second chapter).

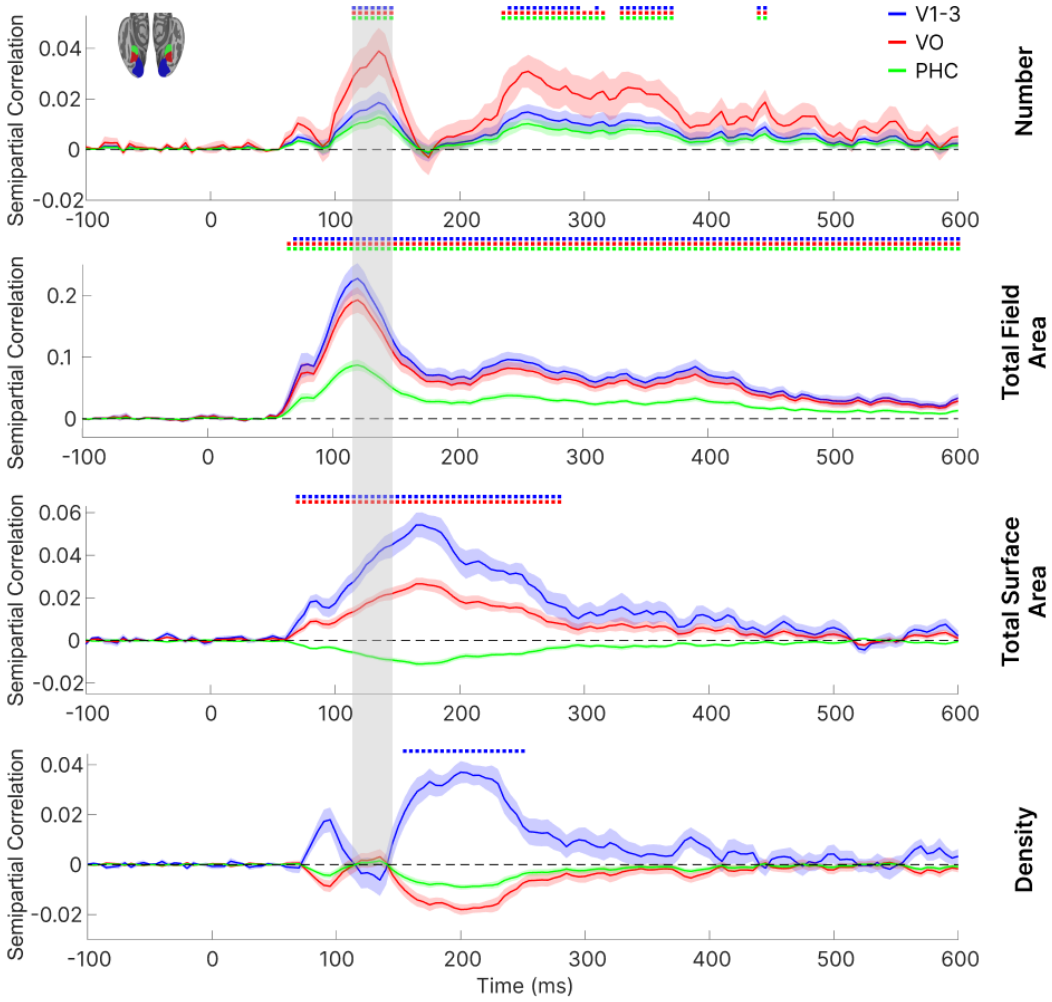



FIGURE 3.9: Time courses of semipartial correlation coefficient depicting the shared variance between model of interest (number, average item area, total field area, total surface area, density), fMRI RDM from the ventral stream and MEG RDM. Standard error of the mean (SEM) across participants is depicted as a shaded area. The horizontal coloured dots indicate significant time points with effects significantly exceeding zero (thresholded at $p < 0.01$, TFCE corrected). The gray rectangle overlaid on each figure highlights that numbers are represented before pairs of non-numeric features (total field area and density or total surface area and average item area) that could indirectly contribute to number computation. It's important to note that the scale of the four represented features differs and the time courses of the average item area are not depicted, as they did not show significance in any region.

## 3.4 Discussion

The current study aimed to explore when numeric and non-numeric features of a visual array of dots are encoded in the brain. In contrast to previous research that examined feature timing by analyzing univariate event-related potential (ERP) signals of EEG (Fornaciai et al., 2017; Park et al., 2015), I employed representational similarity analysis (RSA). This approach enabled us to combine the sensitivity of pattern-based methods with the temporal resolution offered by MEG, providing a more comprehensive understanding of the encoding process.

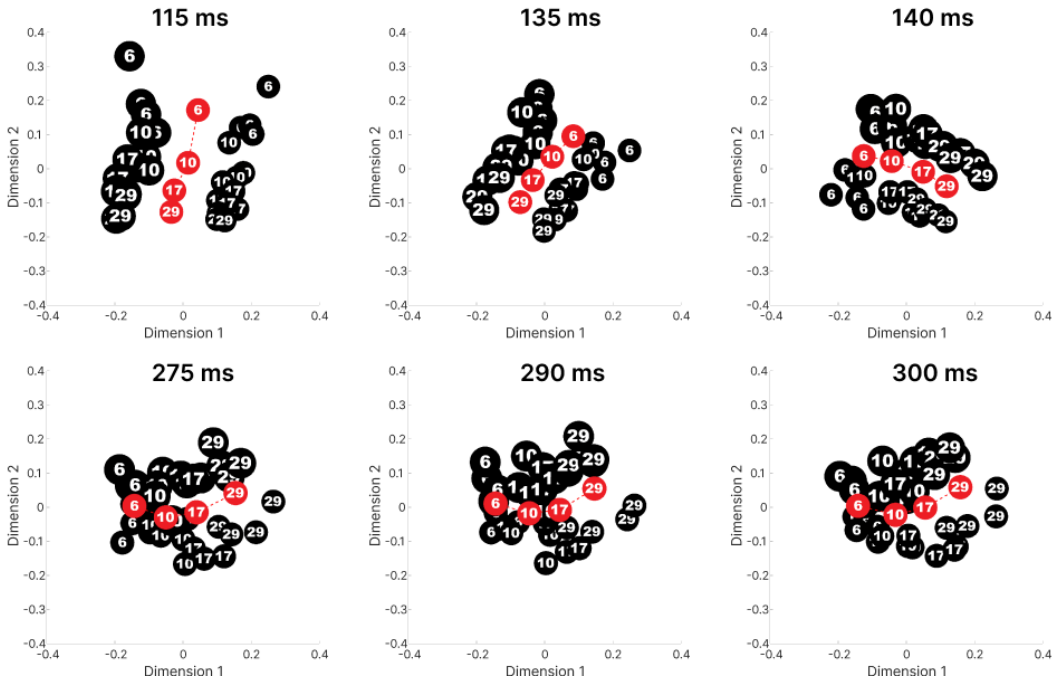



FIGURE 3.10: Multidimensional scaling (MDS) reveals representational similarities between stimuli in a two-dimensional space for three distinct time points from the first period when numerosity becomes significant: 115 ms, 135 ms, and 140 ms and three distinct time points from the later period: 275 ms, 290 ms, and 245-355 ms. The black circles represent the 32 stimuli. The circle sizes vary, indicating stimuli with small total field area (small circles) and larger total field area (large circles). The red circles indicate the average coordinates of each number.

Similar to my previous fMRI study (second chapter), in this study, I employed semi-partial correlation RSA to understand how numeric and non-numeric features within a visual set of dots unfold over time and frequency. Additionally, I extended my analysis by integrating model-based MEG-fMRI fusion techniques, providing a more comprehensive understanding of how numeric and non-numeric features evolve over time across various brain regions. This approach allowed us to investigate whether numerical information emerges hierarchically or simultaneously within the brain and whether it precedes or follows the processing of non-numeric features in different regions.

### Number is a primary visual feature

Using representational similarity analysis I unfolded the temporal dynamics of numeric and non-numeric features of the stimuli. The result suggests that numerosity is represented very early in time, extracted prior to the pairs of non-numeric features from which number can be indirectly computed (see first chapter). This finding suggests that numerosity information is extracted from the visual image through a dedicated mechanism, thus that it can be considered a primary visual property of the image. This finding was further confirmed by the fact that numerosity information

is represented in a different frequency band than the pairs of non-numeric features from which number can be indirectly computed.

To the best of my knowledge, prior to my study, RSA for investigating the timing of numerosity encoding in the brain had only been explored by Bankson et al. (2019). However, their study encountered two significant limitations that complicate the interpretation of their results. Firstly, although they attempted to minimize the influence of non-numeric features on number prediction, they did not account for these effects when correlating neural RDM with number RDM, a step I rigorously followed in this research. Secondly, their selection of numbers did not ensure equal perceptual discriminability, as I did in my study. This raises the possibility that, given my uncertainty about whether these numbers are perceptually discriminable, the observed differences in activation across numbers could be influenced by other non-numeric features that can be perceptually discriminated more easily.

In contrast to my finding that numerosity uniquely explains the variance of MEG signals over other non-numeric features, Gebuis and Reynvoet (2013) reported no numerosity-related effects in the ERP signals of EEG during both passive and active viewing of a visual set of dots. Although they employed a smaller number of subjects, seventeen in total, compared to my study, the doubt in their results arises from the stimulus design, wherein they introduced a non-linear relationship between numerosity and non-numeric features to de-correlate them. This design choice led to a high variance in the non-numeric features. The high variance in non-numeric features, when compared to numerosity, could potentially mask the effect of number in the ERP signals. My findings do fit with two previous studies that used univariate ERP signals in EEG analysis to explore the timing of number representation (Fornaciai et al., 2017; Park et al., 2015). Paralleling my observations, their results indicate a rapid and direct encoding of numerosity information. However, a direct comparison between the timing of my results and these two studies is not straightforward, given that I designed my stimuli and analyzed the data differently. I employed perceptually definable non-numeric features to create my stimuli, while they used two orthogonal mathematically defined constructs (size and spacing) to define theirs (DeWind et al., 2015). Additionally, I ensured that number and average item area were chosen to be equally discriminable and the total field area was selected to allow arrays of dots to be adequately sparse, targeting the number regime rather than the density regime (Anobile et al., 2013), whereas in their case they did not control for the perceptual discriminability of size and spacing since these constructs were mathematical and inherently non-perceptual. In addition to my distinct approach in stimulus design, the current study also differed in data analysis. In their analysis, they only controlled for the effects of size and spacing, leaving open the possibility that the effect of number might reflect a combination of other non-numeric features, such as density and total field area. In contrast, in this study, I simultaneously controlled for the effects of all non-numeric features using semipartial correlation.

Consequently, the effect I reported for numbers reflects its unique influence, with all other non-numeric features having been partialled out.

Neural oscillations are rhythmic patterns of activity in the brain that play a crucial role in how visual information is processed and represented (Ward, 2003). Different frequencies of neural oscillations tend to be associated with different aspects of visual processing. For instance, gamma oscillations tend to be linked to the binding of features into coherent objects (Fries, 2009), while theta oscillations are often involved in spatial navigation (Buzsáki & Moser, 2013). Paralleling previous observations (Rubinsten et al., 2020), I found representation of non-symbolic visual number in the lower beta frequency band. Notably, I observed a shift from the lower beta frequency range to higher beta frequencies in the case of density and total surface area. One plausible explanation for these observations could be that distinct processes underlie the extraction of numerical information as compared to other non-numeric features, i.e. density and total surface area.

**Number is encoded directly in all retinotopically organized regions along the dorsal and ventral stream extremely fast and with no detectable signs of a clear temporal hierarchy**

Previous studies, employing a regression-based approach, have revealed that a significant proportion of the variance in the ERP can be explained by numerosity very early in time, approximately around 75 ms (Park et al., 2015) and 90 ms (Fornaciai et al., 2017) post-stimulus on the medial occipital channel. Based on these findings, it has been concluded that the rapid encoding of numerosity information initiates from the early visual cortex. However, their conclusions are primarily based on the examination of scalp topographies. Relating scalp topographies to underlying source locations is challenging due to the well-established phenomenon in which electrical potentials from various sources blend at the level of scalp EEG recordings (Baillet, 2017). Subsequent studies (Guillaumé et al., 2018; Lucero et al., 2020; Park, 2018; Van Rinsveld et al., 2020) sought to strengthen these initial findings by using steady-state visual evoked potentials (SSVEP). The rationale behind this method is to present stimuli flickering at specific frequencies, thereby tagging the neural activity pattern in the brain with the desired stimulus frequency (Norcia et al., 2015; Vialatte et al., 2010). Due to their robustness against artifacts and high signal-to-noise ratio, these evoked potentials are suitable for isolating responses to numerosity in both children and adults; however, they do not contain temporal information. All of the aforementioned studies, including those conducted by Park (2018) in children and Guillaumé et al. (2018), Lucero et al. (2020), Van Rinsveld et al. (2020) in adults, provide evidence that SSVEPs of the occipital channels of EEG are modulated by numerosity. However, their claim that the early visual cortex is the source of activity

relies solely on the scalp topography of the EEG response, without any source reconstruction to localize the anatomical sources of signals, thereby making it unfounded (Baillet, 2017). Interestingly, Van Rinsveld et al. (2021) recently enhanced the same SSVEP paradigm used by Van Rinsveld et al. (2020) with source localization to identify the anatomical source of evoked potentials. Their findings challenge previous literature that suggested the early visual area as the sole source of numerosity activity based solely on scalp topography results. Instead, they discovered multiple sources of numerosity, including the IPS, supplementary motor area, and middle temporal gyrus.

Here, through the use of model-based MEG-fMRI fusion, I was able to unravel the temporal dynamics of numeric and non-numeric features within various brain regions along both the dorsal and ventral streams, extending beyond the early visual cortex. My findings indicate that numerical information is rapidly encoded along both the dorsal and ventral streams, preceding the pairing of non-numeric features (such as total surface area and average item area, or total field area and density), through which numbers may be indirectly represented by their combination. One possible interpretation of these findings is that, akin to the operation of sensory maps (Young, 1998), where multiple repetitions of topographically organized maps operate parallelly to analyze various aspects of sensory input, different cortical regions in the brain may work in parallel to analyze numerical information. Interestingly, as demonstrated in the first chapter of the thesis using searchlight analysis, the representation of numerosity information is widely distributed across the cortex along both the dorsal and ventral streams, which is comparable to the numerotopic maps discovered by Harvey and Dumoulin (2017a). The distribution of numerosity information across various brain regions, including those associated with different functions such as mathematics, decision-making, motion, and object processing, suggests the use of numerosity information for diverse functions. An alternative interpretation suggests a hierarchical processing of numerosity. In this view, the encoding of numerosity must be abstracted from local spatial representations of visual input to create a global representation, which appears to manifest along the IPS (Viswanathan & Nieder, 2020). Furthermore, Paul et al. (2022) presented evidence that monotonic responses to the number of objects originating in the primary visual cortex (V1) are more closely aligned with aggregate Fourier power than with numerosity. They hypothesized that spatial frequency analyses in early visual processing may underlie neural responses related to numerosity in higher brain regions. They also suggested that a nonlinear interaction among early visual Fourier power responses is essential for converting Fourier power responses into numerosity-tuned responses in the higher cortical areas. While their experimental design lacked conclusive evidence for this interpretation, computational modeling results also indicate a similar mechanism (Kim et al., 2021; Nasr et al., 2019; Stoianov & Zorzi, 2012). It appears that responses in the initial layers of neural networks (Kim et al., 2021; Stoianov & Zorzi, 2012) align more closely with total Fourier power rather than with

numerosity (Paul et al., 2022) and numerosity-tuned units that emerge in the later layers of the neural network result from the weighted sum of units in the earlier layers. Interestingly, in the hierarchical generative network developed by Stoianov and Zorzi (2012), which was trained for efficient image encoding, the deeper layer's numerosity-tuned responses were normalized by the total surface area. This suggests that the total surface area should be presented prior to numerosity which is also consistent with our finding. Nonetheless, the hierarchical representation observed in these neural networks does not necessarily mean that the brain processes information similarly. This is especially evident in a recent computational model that demonstrates numerosity responses can be represented in a single-layer model as opposed to a multi-layer neural network model (Park & Huber, 2022). However, the interpretation that suggests the existence of hierarchical processing is plausible, latency differences between regions remain an open question. For instance, as shown by Nieder and Miller (2004), there are latency differences between the posterior parietal and prefrontal cortex, with the parietal cortex leading significantly by 27 ms. Yet, the scale of latency differences between early visual areas and higher brain regions remains uncertain. If such differences exist at a scale of just a few tens of milliseconds, my analysis approach may not be sufficiently sensitive to capture them.

#### Multiple stages of numerical information processing

The results from representational similarity analysis suggest the existence of at least two and potentially three stages of numerosity processing. The first stage initiates around 115 ms, followed by the second stage around 245 ms post-stimuli and a third one after 300 post-stimulus. These findings parallel previous observations in infants (Gennari et al., 2023) and adults (Fornaciai et al., 2017; Park et al., 2015). The results from adults indicate that the visual cortex seems to process numerical information through two sequential stages, occurring at approximately 100 ms and 150 ms after stimulus onset. One possible explanation of my results is that in the second stage, number is computed through a combination of density and total field area, as both of these features have been encoded prior to the second stage of numerosity representation. However, this explanation appears unlikely, given that the results from temporal generalization analysis suggest a consistent cognitive process underlying number representation over time. This consistency makes it unlikely that numerosity is computed directly at the first stage of representation but it is computed indirectly by combining density and total field area at the second stage. According to an alternative hypothesis, the initial stage of numerosity representation involves an early encoding process which is not sufficient for the conscious experience of number, while the following stages represent the content of the subjective perceptual experience of number (Fornaciai & Park, 2018). Fornaciai and Park (2018) observed

that the ERP signal initially reflects numerosity without any apparent impact of connectedness, with the influence of connectedness becoming evident as early as 150 ms after stimulus onset. According to the connectedness illusion phenomena, connecting items with lines reduces perceived numerosity (He et al., 2009, 2015). Notably, they also identified that the connectedness effect remains robust even in the face of interruptions in feedback signals (Fornaciai & Park, 2021). These results were interpreted as evidence supporting the idea that numerosity representations encompass two distinct stages, with the second stage following the first one through a feedforward segmentation process that underlies the perception of number.

While the current findings do not provide definitive evidence for the conscious experience of number at the second stage and after, as subjects retain the number in their minds without reporting it, it is noteworthy that previous research has indicated that the ERP correlates of visual awareness typically manifest between 170 – 290 ms after stimulus onset (Dembski et al., 2021). Moreover, upon closely examining the MDS results in Video 1, a curved structure emerges at several time points (e.g., 275 ms, 290 ms, 300 ms) post-stimuli. This resembles the structure discovered in both dorsal and ventral associative areas, particularly within the IPS, a region where fMRI decoding performance was associated with behavioral number discrimination acuity (Lasne et al., 2019), as reported in my prior fMRI study (see second chapter). Given this evidence, the suggested explanation that the second stage of numerosity encoding gives rise to the conscious experience of numerosity appears plausible.

## 3.5 Conclusion

In summary, my study employed magnetoencephalography (MEG) to unravel the temporal dynamics of numeric and non-numeric features in the human brain. My findings indicate that numbers are represented at an early stage, preceding pairs of non-numeric features that could indirectly contribute to numerosity computation, suggesting that number is encoded directly and not indirectly computed through the combinations of other features, thus that it is a primary visual feature. Moreover, using model-based MEG-fMRI fusion,I observed rapid representation of numbers along all retinotopic regions of both the dorsal and ventral streams . This occurred before pairs of non-numeric features, from which numbers can be indirectly computed, with no detectable signs of a clear temporal hierarchy.

# CHAPTER 4

## Investigating representation of visual numerosity in machine and its relation to the human brain

# Chapter 4

# Investigating representation of visual numerosity in machine and its relation to the human brain

## 4.1 Introduction

Convolutional neural networks (CNNs) are computational models based on the early discoveries in the study of biological vision (Lindsay, 2021). Similar to the brain, these hierarchical models consist of several feedforward layers. Each layer is composed of numerous artificial units that are meant to simulate neurons in the brain. Since their introduction, they have developed into state-of-the-art models of neural activity and behavior on visual tasks (Cichy, Khosla, et al., 2016; Khaligh-Razavi & Kriegeskorte, 2014; Kubilius et al., 2019; Yamins & DiCarlo, 2016; Yamins et al., 2014). It has been demonstrated that CNNs trained on an object classification task can well account for the brain responses of both humans' and monkeys' inferior temporal cortex (IT), a key region for object recognition (Khaligh-Razavi & Kriegeskorte, 2014). With respect to number processing, it has been recently shown that number-detector units, similar to number neurons recorded in monkey prefrontal and parietal cortex, emerge in a final layer of a CNN trained for visual object recognition (Nasr et al., 2019), and even in a completely untrained CNN (Kim et al., 2021). Notably, the authors discovered that these number-selective units were not influenced by non-numeric visual features. Additionally, Zhou et al. (2021) demonstrated that the phenomenon of numerosity underestimation in connected dots, previously observed in humans (Franconeri et al., 2009; He et al., 2009), also affected the estimation of numerosity in CNNs. These results together with specialized neurons that respond to number of items in visual display in numerically naive monkeys (Viswanathan & Nieder, 2013) and crows (Wagener et al., 2018), untrained 10-day-old domestic chicks (Kobylkov et al., 2022), or neural encoding of number in 3-month-old infants (Gennari et al., 2023) suggest that response to numbers occurs on the basis of processes built into the visual system, which can be simulated by CNNs. While results from previous studies (Kim et al., 2021; Nasr et al., 2019; Zhou et al., 2021) suggest

that a simple feedforward CNN contains numerosity-tuned units, similar to number neurons, it is unknown whether the network represents numbers on a population level and to what extent this representation is similar to the fMRI data of human brains. Here, I investigated how a specific CNN model, CORnet-Z, represents numerical and non-numerical quantities and whether it can capture how the human different brain regions represent numerosity at the population level.

## 4.2 Methods

To test whether CNN models can capture how the human brain represents numerosity over and above the other non-numerical features at the population level, I used CORnet-Z, a model with four anatomically mapped areas (V1, V2, V4, and IT) which is followed by a decoder layer. CORnet-Z is the simplest network in the CORnet family and a lightweight alternative to AlexNet. Each anatomically mapped area in the CORnet-Z consists of a single convolution, followed by a ReLU nonlinearity and max pooling and the decoder is a 1000-way linear classifier (Kubilius et al., 2019). I chose CORnet-Z because it balanced the resemblance to the architecture that was used by previous studies on numerosity and because it well fit visual system. I used the three versions of the CORnet-Z:

1. the completely untrained version with randomly initialized weights to reveal the effect of architecture alone (Cichy, Khosla, et al., 2016)(Cichy et al., 2016),

2. a version trained on object recognition using the ImageNet dataset (Deng et al., 2009), which contained 1.2 million images of objects over 1,000 categories (Krizhevsky et al., 2012), as it has been used in a previous study by Nasr et al. (2019), and

3. a version trained on SOS The Salient Object Subitizing dataset (SOS; J. Zhang et al., 2015) which contains about 14,000 everyday images sourced from four widely used image datasets: COCO (Lin et al., 2014), ImageNet (Deng et al., 2009), VOC07 (Everingham et al., 2009), and SUN (Xiao et al., 2010). The SOS dataset contains five different categories: 1. category zero: images without any salient objects, 2. category one: images with only one salient object, 3. category two: images with two salient objects, 4. category three: images with three salient objects, 5. category four and more: images with four or more than four salient objects (Figure 4.1). Given that over two-thirds of the images in the aforementioned datasets belong to the zero or one category, a cut-and-paste approach was employed to generate synthetic images for categories lacking sufficient data. For these synthetic images, salient objects were selected from the THUS10000 dataset (Cheng et al., 2015). I opted to use this version of CORnet-Z to explore the effect of augmenting the network with the task to recognize the number of objects (subitizing task). I fine-tuned a pre-trained CORnet-Z

model on the ImageNet dataset using the SOS dataset for 50 epochs. The training employed Stochastic Gradient Descent (SGD) with a learning rate of 0.001. The training process was implemented using a custom PyTorch code as implemented in Mistry et al. (2023) (available at https://github.com/scsnl/Mistry_Strock_NatureComm_2023). The network's subitizing performance increased to 65% after fine-tuning on the SOS dataset.


FIGURE 4.1: Sample images from SOS dataset (figure adapted from J. Zhang et al. (2015)).

After training, the three versions of the model were fed with 100 images for each of the 32 conditions (4 numerosities, 4 average item areas, and 2 total field areas) of the same dots images used in the fMRI and the MEG studies presented in chapter one and two of this thesis. An image with a size of 500 × 500 pixels served as the input to the network. I selected four layers (V1, V2, V4, and IT) of the network which are analogous to visual brain areas and extracted the activation of all nodes in each of the layers. I extracted the activation using the THINGSvision toolbox (Muttenthaler & Hebart, 2021) and averaged the results of the 100 instances of each condition to get one vector of activity for each condition from each layer' output. Similar to the process of making neural RDM from selected voxels activation, I used the Pearson correlation to build the CORnet-Z's RDMs. The model RDMs for each feature and neural RDMs were computed as described in previous chapters.

### 4.2.1 Comparing convolutional neural network with predictor models

To investigate whether the representations of numerical and non-numerical features of the stimuli can be disentangled from the layers of the networks I employed representational similarity analysis combined with semipartial correlation. To assess the significance of the resulting semipartial correlation values I performed statistical inference on the correlation using permutation test (Nili et al., 2014). To do this, I first shuffled the stimulus labels and then I calculated the RDM of CORnet-Z layers based on the permuted labels stimulus. I repeat this process for 50000 times to generate a null distribution of semipartial correlation. I reject the null hypothesis of unrelatedness of dissimilarity matrices if the actual semipartial correlation is inside the top 5% of the simulated null distribution of semipartial correlations with a false-positives rate of 0.05 (Kriegeskorte et al., 2008).

### 4.2.2 Comparing convolutional neural network to human fMRI data

To address the question of how much variance in human fMRI brain data can be explained by the layers of the networks, I examined the correlation between different layers of the networks and the human fMRI data using Pearson correlation. The correlation coefficients were then normalized by the noise ceiling of the corresponding region and then one-sample t-tests against zero across subjects were used to test the statistical significance of correlation coefficients for each ROI.

### 4.2.3 Controlling for the convolutional neural network features

To determine whether numerosity representation in the human fMRI data is similar to the representation of number by the networks, I performed a semipartial correlation analysis between neural RDM and number RDM while partialling out all non-numeric features RDMs and RDMs obtained from layers of the networks (Figure 4.2). The semipartial correlation coefficients were then normalized by the noise ceiling of the corresponding region and one-sample t-tests against zero were used to test the statistical significance of correlation coefficients for each ROI.

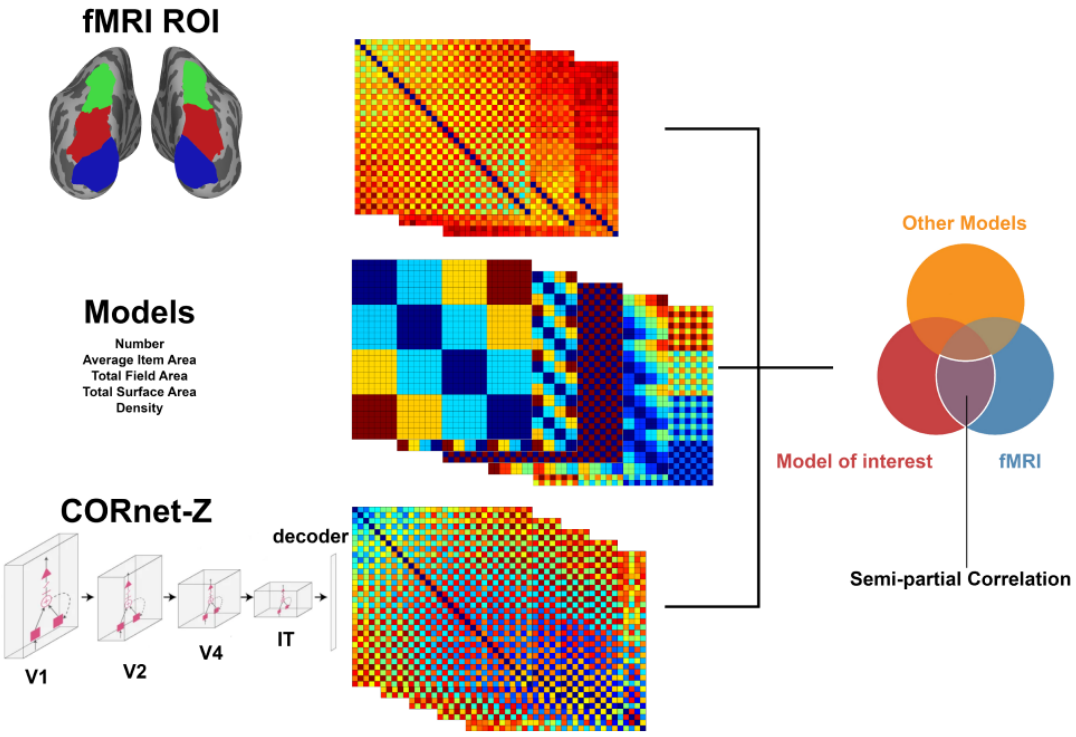



FIGURE 4.2: Neural representational dissimilarity matrices (RDM) obtained from fMRI data were subjected to semipartial correlation analysis. Number RDM were used as a model of interest in the semipartial correlation analysis while the effect of all non-numeric model RDMs and RDMs derived from CORnet-Z layers were partialled out.

## 4.3 Results

### 4.3.1 Results of representational similarity analysis on CNN layers with features as predictors

Similar to the analysis I did with the fMRI data from different regions of interest, I used representational similarity analysis to disentangle the contribution of numerical and non-numerical features of the stimuli on the CORnet-Z layers patterns of activity. Figure 4.3 shows the results from semipartial correlation between the RDMs derived from CORnet-Z and model RDMs across different layers of the untrained network and the network trained on ImageNet dataset, respectively. The result shows that the variance in layers activations is significantly explained by number over and above the other features already in deeper layers of the neural network (Untrained network: IT layers, Network trained on ImageNet: V4 and IT, $p<0.05$). I detect a similar pattern for average item area in the CORnet-Z trained on ImageNet, total surface area in the untrained CORnet-Z, and total surface area in both versions of the network while the opposite pattern of results was true for density in both untrained network and the network trained on ImageNet: density was higher at earlier layers while it gets lower at the deeper layers.

I replicate the same analysis for CORnet-Z fine-tuned on SOS. Figure 4.3 depicts the results from semipartial correlation between the RDMs derived from CORnet-Z and model RDMs across different layers of the network trained on ImageNet dataset and the network fine-tuned on SOS dataset, respectively. No discernible differences appear to be present in the four layers of the network.

### 4.3.2 Results of representational similarity analysis on fMRI data with CNN layers as predictors

I then quantified the degree of similarity between the CORnet-Z layers and different predefined retinotopic brain regions across the dorsal and ventral streams (L. Wang et al., 2014) using representational similarity analysis. The results revealed that in both the untrained network and the network trained on ImageNet, all layers of the network exhibited high correlations between the fMRI data and the network's layers. These correlations gradually decreased as I moved closer to the higher brain areas in the dorsal and ventral streams (Figure 4.4).

Next, I measured the extent of similarity between the layers of CORnet-Z fine-tuned on SOS and predefined retinotopic brain regions (Figure 4.4). Once again, no discernible differences were observed between CORnet-Z trained on ImageNet and CORnet-Z fine-tuned on SOS.

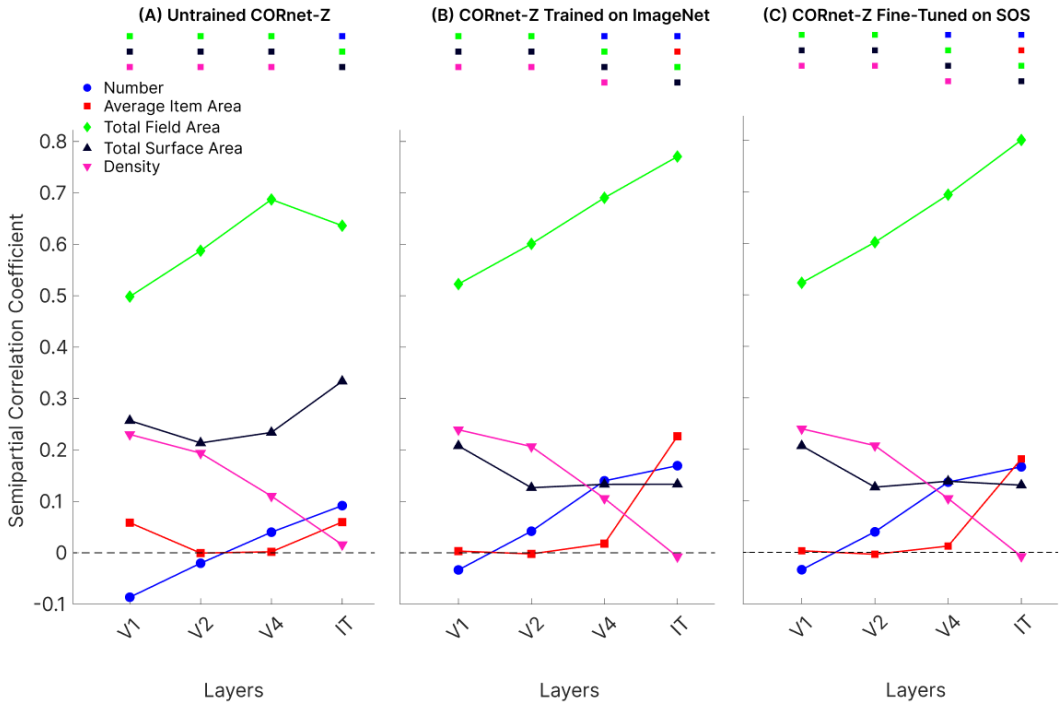



FIGURE 4.3: Semipartial correlation coefficient obtained from the representational similarity analysis for number, average item area, total field area, total surface area and density from different layers of (A) untrained CORnet-Z, (B) CORnet-Z trained on ImageNet, and (C) CORnet-Z fine-tuned on SOS. While all non-numeric features, except average item area, is represented in almost all layers of CORnet-Z, number is represented in the last layers of CORnet-Z (untrain CORnet-Z: layer IT; CORnet-Z trained on ImageNet: layer V4 and IT; and CORnet-Z fine-tuned on SOS: layer V4 and IT). Data points show mean semipartial correlation coefficient. The coloured points above the figure indicate where the effect significantly exceeds zero ($p<0.05$). Semipartial correlation coefficients were not normalized here by a noise ceiling due to the absence of neuroimaging data.

### 4.3.3 Results of representational similarity analysis on fMRI data with CNN layers and features as predictors

I then quantified the similarity between the number model's RDM and the RDMs extracted from fMRI data, while controlling for the influence of all non-numeric model RDMs and the RDM extracted from the IT layer of CORnet-Z. This analysis aimed to determine whether CORnet-Z's IT layer could account for the numerical information present in various brain regions across both the dorsal and ventral streams. I focus my report solely on the IT layer of CORnet-Z because it was the only layer containing numerosity information in both untrained CORnet-Z and CORnet-Z trained on ImageNet, as indicated by previous analyses. Additionally, the IT layer exhibited a high correlation with brain regions, as established in the preceding section. The regions of interest analysis shows that after removing CORnet-Z's IT layer numerosity information became non-significant in the early visual areas with reduced number information (all brain regions $p<0.001$) but it remained significant in the higher-level areas along both the dorsal, up to parietal regions, and ventral stream,

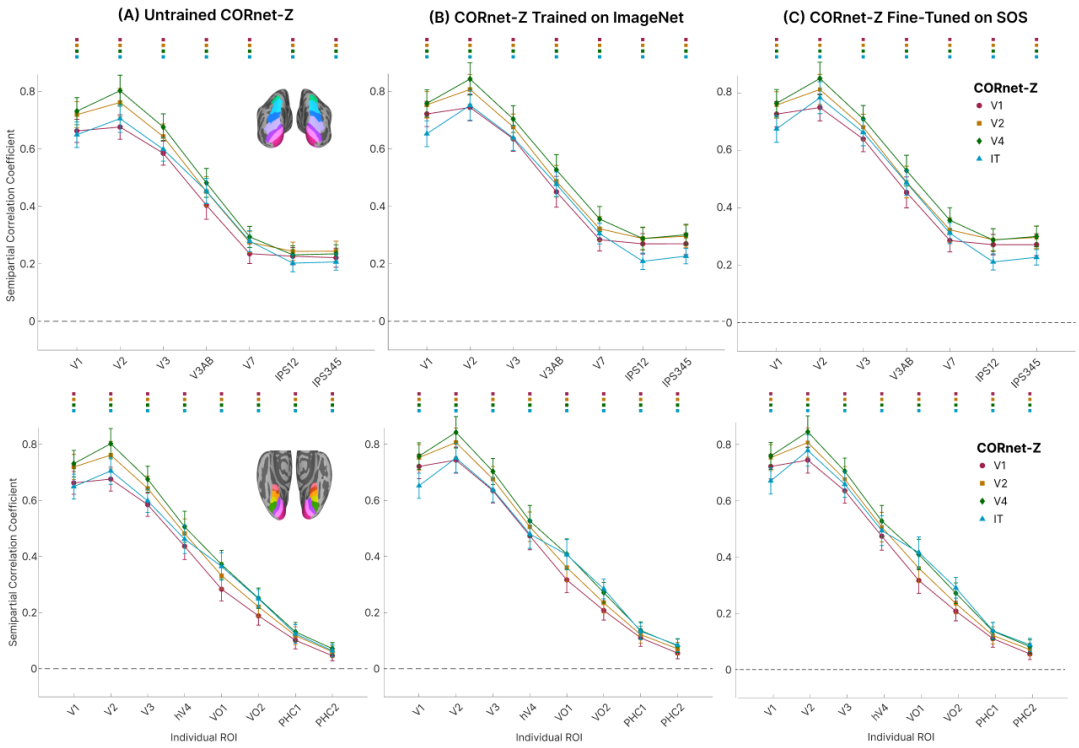



FIGURE 4.4: Semipartial correlation coefficient obtained from the representational similarity analysis for five layers of CORnet-Z (V1, V2, V4, and IT) from predefined dorsal and ventral retinotopic ROIs as defined in the probabilistic atlas by L. Wang et al. (2014) for (A) untrained CORnet-Z and (B) CORnet-Z trained on ImageNet, and (C) CORnet-Z fine-tuned on SOS. Data points show mean semipartial correlation coefficient across subjects (n = 31) ± standard error of the mean (SEM). The coloured points above the figure indicate where the effect significantly exceeds zero ($p<0.01$). All the semipartial correlation coefficients were normalized according to the noise ceiling of their corresponding regions, as detailed in the supplementary materials.

up to occipitotemporal regions (Figure 4.5), suggesting that CORnet-Z' IT captures well the way in which early visual regions represent numerosity but not so well the representational geometry of higher-level ones.

Then, to statistically support the differential impact of the variables of interest across the ROIs, I conducted analyses on semipartial correlation coefficients using four repeated measures ANOVAs (as described below). In these analyses, ROIs and the variables of interest were treated as factors. The variables of interest, in this context, refer to: (I) the resulting semipartial correlation coefficients between number RDM and fMRI RDM while partialling out the effect of other non-numeric features and (II) the resulting semipartial correlation coefficients between number RDM and fMRI RDM while partialling out the effect of other non-numeric features and the RDM extracted from IT layer of CORnet-Z trained on ImageNet. ROIs were considered both aggregated in three big regions (not depicted in the Figure, see the second chapter), and separately for each of the individual regions within the ventral and dorsal streams. The significant two-way interaction between ROIs and the variables of interest, observed across all four ANOVAs, confirmed that the semipartial correlation

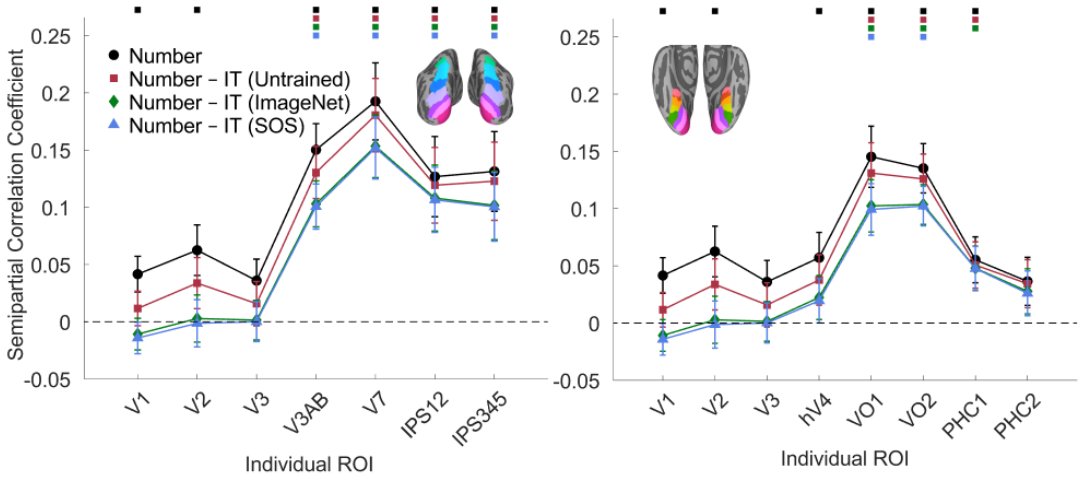



FIGURE 4.5: Semipartial correlation coefficient obtained from the representational similarity analysis between predefined dorsal and ventral retinotopic ROIs and number model RDM. This analysis controlled for the influence of non-numeric number model RDM (represented by the black line), non-numeric model RDM and the IT layer from the untrained CORnet-Z (indicated by the red line), as well as non-numeric model RDM and the IT layer from CORnet-Z trained on ImageNet (depicted by the green line) and CORnet-Z fine-tuned on SOS (depicted by the blue line). Data points show mean semipartial correlation coefficient across subjects (n = 31) ± standard error of the mean (SEM). The coloured points above the figure indicate where the effect significantly exceeds zero ($p<0.01$).

coefficient estimated for the different variables of interest were different across ROIs in both the dorsal stream hierarchy (for the three large regions: $F(1.82,54.46)=19.238$, $p<0.001$; for the individual regions: $F(1.21,36.16)=16.022$, $p<0.001$) and in the ventral stream hierarchy (for the three large regions: $F(1.82,54.53)=31.797$, $p<0.001$; for the individual regions: $F(1.30,38.90)=14.891$, $p<0.001$). I then replicated the aforementioned analysis using two different variables of interest: (I) the resulting semipartial correlation coefficients between number RDM and fMRI RDM while partialling out the effect of other non-numeric features and (II) the resulting semipartial correlation coefficients between number RDM and fMRI RDM while partialling out the effect of other non-numeric features and the RDM extracted from IT layer of untrained CORnet-Z. Once more, the significant two-way interaction between ROIs and the variables of interest, identified in all four ANOVAs, reaffirmed the distinctiveness of semipartial correlation coefficients calculated for the variables of interest across ROIs. This pattern held true within both the dorsal stream hierarchy (for the three large regions: $F(1.94,58.18)=28.311$, $p<0.001$; for the individual regions: $F(1.17,35.19)=11.265$, $p=0.001$) and the ventral stream hierarchy (for the three large regions: $F(1.63,49.05)=41.538$, $p<0.001$; for the individual regions: $F(1.30,38.97)=11.174$, $p=0.001$).

I performed the same representational similarity analysis as outlined above for the IT layer of CORnet-Z fine-tuned on SOS (Figure 4.5). Once again, no discernable differences were detected between CORnet-Z trained on ImageNet and CORnet-Z

fine-tuned on SOS.

### 4.3.4 Results of applying multidimensional scaling on layers of CNN

Finally, to further investigate the latent similarity structure of the CORnet-Z's layers—specifically V1, V2, V4, and IT—in a data-driven manner, I employed multidimensional scaling. This was carried out in a similar way that has been outlined in previous chapters. MDS was applied on the dissimilarity matrix extracted from the four layers of the three networks (untrained CORnet-Z, CORnet-Z trained on ImageNet, CORnet-Z fine-tuned on SOS). The results from the MDS (see Figure 4.6) reveal a rank-ordering of numbers along the second dimension across all four layers of the networks. Additionally, there is a distinct separation between stimuli with large and small total field area apparent in the results for all layers of the networks. This implies a difference in the representation of stimuli within the network compared to higher associative areas of the brain. Notably, in the MDS of higher associative areas of both ventral and dorsal stream, this separation is less pronounced (see chapter two).

## 4.4 Discussion

Together with that of others my work has demonstrated that numerosity is represented in the brain in several areas all along the dorsal and the ventral stream, starting from early visual areas and up to parietal and occipitotemporal regions (see second chapter). Intensive modeling efforts have been made to explore how a computational system can effectively represent numerosity information when presented with visual images (Dakin et al., 2011; Dehaene & Changeux, 1993; Hannagan et al., 2018; Kluth & Zetzsche, 2016; Knops et al., 2014; Park & Huber, 2022; Paul et al., 2022; Stoianov & Zorzi, 2012; Testolin et al., 2020; Verguts & Fias, 2004).

Testolin et al. (2020) employed a similar approach, using RSA along with dimensionality reduction for 2D visualization through t-SNE, to assess visual sense of number in deep belief networks. These networks are capable of deriving complex internal representations of the environment from sensory data (Zorzi et al., 2013). While my approach involved training the CORnet-Z network through supervised learning, the approach taken by Testolin et al. (2020) involved unsupervised learning. Unsupervised learning is often preferred over supervised learning for its biological and psychological plausibility (D. D. Cox & Dean, 2014; Zhuang et al., 2021), reflecting real-world scenarios where human infants learn without explicit millions of category labels (Bergelson & Swingley, 2012; Frank et al., 2021). Despite this, supervised models have shown a closer alignment with the representations found in the ventral cortex (Khaligh-Razavi & Kriegeskorte, 2014). Comparing my results directly with those of Testolin et al. (2020) is not straightforward due to differences in training methodologies and datasets. My network was trained using ImageNet and the SOS

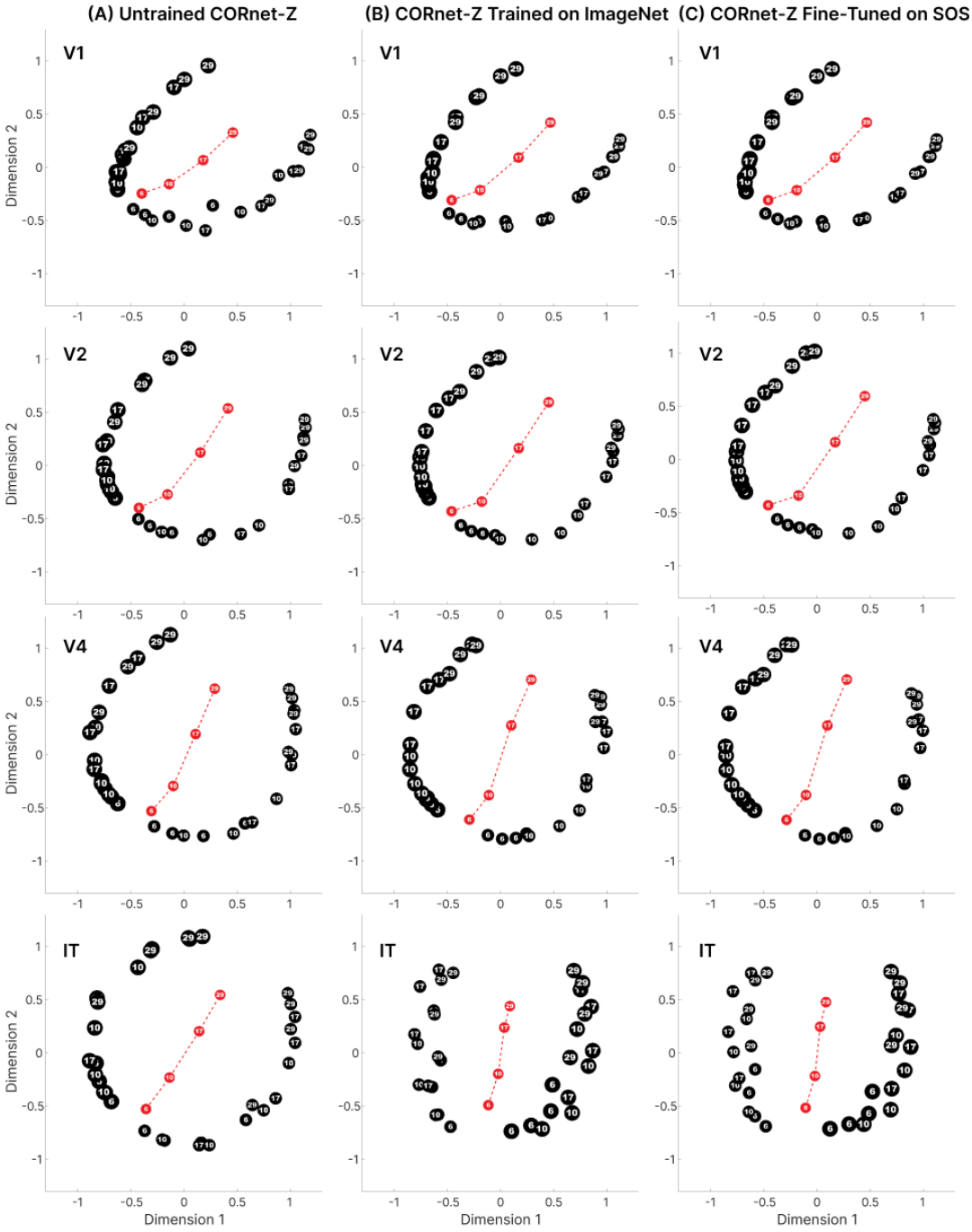



FIGURE 4.6: Multidimensional scaling (MDS) reveals representational similarities between stimuli in a two-dimensional space for four layers (V1, V2, V4, and IT) of (A) untrained CORnet-Z , (B) CORnet-Z trained on ImageNet, and (C) CORnet-Z fine-tuned on SOS. The black circles represent the 32 stimuli. The circle sizes vary, indicating stimuli with small total field area (small circles) and larger total field area (large circles). The red circles indicate the average coordinates of each number.

dataset, whereas their network was trained using artificial images of dot sets, similar yet not identical to the stimuli used in our fMRI experiments. Upon a closed examination of the results, both similarities and differences emerge. In both Testolin et al.

(2020) and CORnet-Z, numerosity information is represented along other visual features. In the network by Testolin et al. (2020), numerosity becomes the most salient feature after training, aligning with the hypothesis that humans can learn to focus more on number and ignore irrelevant non-numerical features (Piazza et al., 2018). Conversely, in my network, the total field area was more salient than numerosity, as it was Testolin et al. (2020) young network. However, training the CorNet-Z with ImageNet and the SOS dataset improved its ability to represent numerosity, as shown by higher semipartial correlation values in the trained vs. untrained versions. This finding would fit with fMRI studies showing sharper numerosity tuning functions in adults (Piazza et al., 2004) compared to preschoolers (Kersey & Cantlon, 2016).

Despite significant efforts to model numerosity information extraction from visual sets of dots, there is no research comparing these models with population-level neural data. As state-of-the-art models of neural activity on visual tasks, CNNs have been shown to encode the number of objects through single units tuned to specific numerosities (Kim et al., 2021; Nasr et al., 2019). However, this research has been criticized, both for the statistical analyses adopted to identify such few number-sensing units over a large set of units, and for the relatively small number of sample images used for testing them (X. Zhang & Wu, 2020). To address these challenges I have taken two key steps. Firstly, I significantly increased the size of my test dataset by using 3200 sample images, a notable expansion compared to the 336 sample images used in previous studies. Secondly, I adopt a distinct approach in my data analysis, going beyond the search for single numerosity-tuned units and examining the population level response. Furthermore, I directly explore the extent to which the polulation-level representations resemble that of the human brain as assessed through the model-based RSA analysis of the fMRI BOLD signal. To accomplish this aim, I quantified the representation of numerosity across many units in the CNN' layers using model-based RSA and compared it to the neural model-based RSA results. Moreover, I explore the representational geometries of different layers to that of different brain regions. My analysis demonstrated

1. that the last layer (layer IT) of all three versions of CORnet-Z are capable of encoding numerical information.

2. that in all three version of CORnet-Z all layers of the network, except the decoder layer, exhibited high correlations with the fMRI data. These correlations gradually decrease as I moved closer to the higher brain areas in the dorsal and ventral streams, suggesting that the current models still fall considerably short of being acceptable models of higher brain areas in both the dorsal and ventral streams for the stimulus space. This result is further confirmed by examining the latent similarity structure of all layers of the CNN on the stimulus set using MDS. Despite the diminished difference between stimuli with large and small total field area in higher brain regions (specially IPS), a distinct separation between them is evident in all layers of the network.

3. that after accounting for the effect of CORnet-Z's IT layer, the fMRI activation in early visual area no longer exhibits numerosity information, suggesting that the network extremely well simulates how the human early visual regions represent numerosity. However, the CORnet-Z failed to explain the numerosity information observed in the fMRI signal from higher cortical areas of both the dorsal and the ventral regions. This highlights that the models do not represent numerosity information in similar ways as the higher-level regions of the human brain do.

**Augmenting the CNN training with a subitizing task does not reduce the gap between the model and brain**

It has been suggested that there are two main anatomically distinct pathways in the brain: the dorsal, or "where", and the ventral, or "what", pathway (Goodale & Milner, 1992; Mishkin et al., 1983). The functional role of each pathway also seems different. The dorsal pathway is though as coding for abstract relation among the objects, spatial position and saliency of of object in a scene (Bisley & Goldberg, 2010; Itti & Koch, 2001; Summerfield et al., 2020) while the ventral pathway is specialized for visual object recognition (Bracci & De Beeck, 2023; Gauthier & Tarr, 2016). A CNN trained in a supervised manner to categorize objects has demonstrated the ability to simulate object representations in the ventral stream (Khaligh-Razavi & Kriegeskorte, 2014; Storrs et al., 2021; Yamins et al., 2014). However, fewer studies have explored the similarity of dorsal representations with deep networks (Bakhtiari et al., 2021; Güçlü & Van Gerven, 2017). According to classical connectionist models of numerosity perception, objects are mapped onto a location map during the process of extracting numerosity from visual images (see first chapter for more detail). This concept of a location map in their model may be connected to the saliency map, which has been proposed to reside in the parietal cortex (Roggeman et al., 2010). While some experimental evidence suggests that our ability to enumerate small groups of stimuli is based on a salience map (e.g. Fu et al., 2022; Knops et al., 2014; Melcher & Piazza, 2011; Piazza et al., 2011), which represents the locations of important items, there is ongoing debate regarding whether this saliency map architecture may also apply to number estimation of large sets (Castaldi et al., 2021; Roggeman et al., 2010). The proposed computational mechanism of the saliency map consists of a single layer of an on-center off-surround recurrent network. In this network, each neuron has an excitatory connection with itself and inhibitory connections with all other neurons, with each node representing a specific spatial location. Higher levels of lateral inhibition result in more precise representations, yet they also limit the capacity of the map. Conversely, lower levels of lateral inhibition lead to coarser representations, and a higher capacity limit (Knops et al., 2014; Roggeman et al., 2010; Sengupta et al., 2014; Verma & Sengupta, 2023). Predictions from these models have been corroborated by neurophysiological data from monkeys (Bisley & Goldberg, 2003) and fMRI data from human subjects (Knops et al., 2014; Roggeman

et al., 2010). Roggeman et al. (2010) found that during a numerosity estimation task with numbers exceeding the subitizing range, activity in the IPS regions rose for up to eight items before slightly decreasing, consistent with salience map models featuring medium lateral inhibition settings. In contrast, when participants undertook a similar but less challenging task, IPS activity initially decreased from four to eight or sixteen items, then increased again, mirroring the pattern seen in models with low inhibition settings. This implies that the on-center off-surround recurrent network architecture, through adjusting the degree of lateral inhibition, may underlie processing both small and large numbers (Sengupta et al., 2014; Verma & Sengupta, 2023). Moreover, emergent evidence suggests CNNs trained on ImageNet may use texture and context (Baker et al., 2018; Gatys et al., 2017; Geirhos et al., 2018) for successful object classification instead of explicitly representing object shapes (Kubilius et al., 2016). Local differences in texture within an image may be what the networks trained with ImageNet could pick up in terms of saliency and thus use it for estimating numerosity. Consequently, training on the SOS dataset could help mitigate this bias. However, for the moment this is simply a speculation that future studies could maybe further address. By drawing inspiration from these previous studies, I tried to augment CNN training procedures by incorporating salient objects detection task. Specifically, I trained CORnet-Z to subitize salient objects using the SOS dataset, aiming to minimize the gap between CORnet-Z results and fMRI data derived from the dorsal stream. This approach was motivated by previous findings indicating that face-processing signatures are present in networks trained on face recognition but not in those trained on object recognition (Dobs et al., 2023; Kanwisher et al., 2023). However, augmenting the network training procedure does not appear to impact its performance on the numerosity detection task, as indicated by the results from RSA and MDS analysis with the different layers of networks. One possible explanation for this result is that the network has been trained to detect a range of numbers within the subitizing range (see the first chapter), while my stimuli extend beyond this range and it has been suggested that two distinct neural mechanisms underlie the representation of small and large numbers (Kutter et al., 2023; Revkin et al., 2008). Therefore, it is essential to train a network with a range of numbers beyond subitizing. This need may be further supported by recent findings from Mistry et al. (2023), who observed that after training a CNN for a numerosity classification task, a curved structure resembling the pattern discovered in the parietal region in the second chapter of the thesis emerges in the last layer of the network.

## 4.5 Conclusion

In summary, my study revealed that, similarly as humans, the final layers of both untrained CORnet-Z and CORnet-Z trained on ImageNet to discriminate images and on SOS to enumerate salient objects equally encode numerosity information beyond the non-numeric features of visual sets of dots. However, while the deep layers of

the network displayed high correlation with fMRI data from early visual areas of human subjects, it falls short in explaining the variance of numerosity representation as it progresses to higher brain regions along both the dorsal and ventral streams. Furthermore, training the network to subitize salient objects did not alleviate the gap between the network and the brain.

# CHAPTER 5

## Conclusion

# Chapter 5

# Conclusion

## 5.1 Summary

In my thesis, I presented evidence regarding where and when numeric information extracted from a visual set of dots is represented in the human brain. Through the unique analytical approach I employed, I successfully separated numerical information from non-numerical features within the visual set of dots. This disentanglement allowed me to discern how each feature is distinctly represented in the brain. Additionally, I could integrate the results from different modalities, fMRI and MEG, facilitating the identification of the spatio-temporal representation of each feature in the brain. Also, I compared the fMRI result with a computational model, a CNN, to identify both the similarities and differences in the representation of visual sets of dots in the CNN and the human brain. I began this work by asking three questions, each shaping the focus of a specific chapter in my thesis. Now, I am able to provide brief answers to each question, thereby concluding the thesis.

**Question 1. Do brain regions aside from the parietal region also represent pure numerical information when subjects are presented with visual sets of dots?**

The answer is YES. Prior research has provided evidence indicating the presence of numerical information in the parietal, frontal, and early visual regions of the brain. With this study I was able to provide evidence suggesting that numerical information representation extends to brain regions beyond the aforementioned areas, notably encompassing regions within the ventral stream. Moreover, I also showed that associative regions of both streams encode numerosity in a very similar way, and differently from primary visual regions, indicating that the former and not the latter might be involved in numerical decision-making.

**Question 2. When numeric and non-numeric features of a visual set of dots are represented in the brain?**

I addressed this question through the application of representational similarity analysis. My findings indicate that the total field area and total surface area of a visual set of dots are represented prior to the numerical information. However, number is represented before density, very early around 115 ms after presenting the stimulus to participants. Furthermore, my analysis did not reveal any evidence of the representation of average item area over time. Consequently, it can be deduced that number number is represented before pairs of non-numeric features, such as the total field area and density or total surface area and average item area. This implies that number is a primary visual feature, directly encoded rather than computed through the combination of other non-numeric features. Next, by expanding my analysis and integrating the MEG results with the fMRI findings, I discovered that numerical information is encoded extremely fast along both the dorsal and ventral pathways.

**Question 3. To what extent do the representations of numerical information in the Convolutional Neural Network (CNN) align with the fMRI data of the human brain at the population level?**

I answered this question by employing representational similarity analysis to examine the capacity of a CNN to represent number at the population level. My analysis revealed that, firstly, the CNN, specifically CORnet-Z, whether untrained or trained on ImageNet, is capable of representing numerical information. However, the manner in which CORnet-Z represents numbers differs from the representation of numbers in higher brain regions. It does, however, exhibit the ability to simulate the representation of numbers in the early visual areas (V1 – V3).

## 5.2 Future Work

The work presented in this study aimed to address specific questions as outlined earlier. Nonetheless, there exist several research directions that I would like to highlight for future investigations, outlined below.

### 5.2.1 The ventral stream representation of numerosity

In this study, I could find regions within the ventral stream, which I have referred to as NTO due to its apparent overlap with a numerotopic map previously reported by Harvey and Dumoulin (2017a). Nevertheless, the functional significance of this numerosity map and the mechanisms governing its formation remain debated. It remains unclear whether this region serves a distinct functional role compared to the established number form area (NFA). This area might emerge as a result of functional connectivity between the dorsal and ventral regions, possibly influenced by

formal math education. Alternatively, the ventral stream could independently extract numerosity information, a possibility not necessarily exclusive with the previous explanation. Future works can differentiate between these two explanations using functional connectivity analysis and by conducting experiments with preschool children or innumerate adults.

### 5.2.2 Computational model of numerosity extraction from visual set of dots

Despite extensive research on computational models for extracting numerosity information from visual dot sets, the precise mechanism by which the brain accomplishes this remains unknown. While my analysis indicates that CCNs could be potential candidates for numerosity extraction in early visual areas, the mechanisms underlying numerosity extraction in higher brain areas are still unclear. One potential research direction for future modeling could involve training CNNs with tasks other than object recognition (Kanwisher et al., 2023) to investigate their ability to explain numerosity information in higher brain regions.

### 5.2.3 Extraction of numerosity information from auditory stimulus

Although one conventional approach to assess numerosity perception is through the visual modality, where all stimuli are presented simultaneously, it's important to note that this is not the exclusive method for evaluating it. In a previous study with monkeys (Nieder et al., 2006), examining both sequential and simultaneous presentations of visual sets of dots, the findings suggested that distinct populations of neurons were involved in the ongoing quantification processing for both presentation methods. However, the final result was encoded by a different population of neurons, irrespective of how the stimulus was presented. A recent study with infants suggests that numbers are encoded in the infant brain in a supramodal manner (Gennari et al., 2023). Employing EEG and sequential auditory stimuli, the study demonstrated the spontaneous encoding of numbers in the brain. Notably, they controlled for non-numeric auditory features, specifically tone rate and total sequence duration, which can be likened to density and total field area when testing with a visual set of dots. For future investigations, it would be worthwhile to explore which brain areas encode both numeric and non-numeric auditory sequences of tones. This exploration can help determine whether the neural code for numbers generalizes from a visual set of dots to auditory sequences of tones or not.

### 5.2.4 Numerosity representation in the natural context

Almost all studies exploring numerosity representation in both humans and animals have relied on artificial stimuli, primarily visual arrays of dots. However, in natural environments, we encounter complex objects. It is crucial to investigate the distribution of numeric and non-numeric features of these objects in their natural

context and understand how this distribution might influence numerosity estimation. If there exists a bias in the natural environment, such as a correlation between the number of objects and their total surface area, it is essential to examine how this bias impacts numerosity estimation and in what manner.

# Bibliography


Abdi, H. (2007). Part and partial correlations. SAGE Publications.

Acharya, D., Rani, A., Agarwal, S., & Singh, V. (2016). Application of adaptive Savitzky–Golay filter for EEG signal processing. *Perspectives in science*, *8*, 677–679. https://doi.org/10.1016/j.pisc.2016.06.056

Agrillo, C., & Bisazza, A. (2018). Understanding the origin of number sense: a review of fish studies. *Philosophical Transactions of the Royal Society B*, *373*(1740), 20160511. https://doi.org/10.1098/rstb.2016.0511

Agrillo, C., Petrazzini, M. E. M., Tagliapietra, C., & Bisazza, A. (2012). Inter-Specific differences in numerical abilities among teleost fish. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00483

Al-Tahan, H., & Mohsenzadeh, Y. (2021). Reconstructing feedback representations in the ventral visual pathway with a generative adversarial autoencoder. *PLOS Computational Biology*, *17*(3), e1008775. https://doi.org/10.1371/journal.pcbi.1008775

Amalric, M., & Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(18), 4909–4917. https://doi.org/10.1073/pnas.1603205113

Anobile, G., Cicchini, G. M., & Burr, D. (2013). Separate mechanisms for perception of numerosity and density. *Psychological Science*, *25*(1), 265–270. https://doi.org/10.1177/0956797613501520

Anobile, G., Cicchini, G. M., & Burr, D. (2015). Number As a Primary Perceptual Attribute: A Review. *Perception*, *45*(1-2), 5–31. https://doi.org/10.1177/0301006615602599

Ansari, D., & Dhital, B. (2006). Age-related Changes in the Activation of the Intraparietal Sulcus during Nonsymbolic Magnitude Processing: An Event-related Functional Magnetic Resonance Imaging Study. *Journal of Cognitive Neuroscience*, *18*(11), 1820–1828. https://doi.org/10.1162/jocn.2006.18.11.1820

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*(3), 695. https://doi.org/10.2307/1130057

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Arrighi, R., Togoli, I., & Burr, D. (2014). A generalized sense of number. *Proceedings of The Royal Society B: Biological Sciences*, *281*(1797), 20141791. https://doi.org/10.1098/rspb.2014.1791

Ayzenberg, V., & Behrmann, M. (2022). Does the brain's ventral visual pathway compute object shape? *Trends in Cognitive Sciences*, *26*(12), 1119–1132. https://doi.org/10.1016/j.tics.2022.09.019

Baillet, S. (2017). Magnetoencephalography for brain electrophysiology and imaging. *Nature Neuroscience*, *20*(3), 327–339. https://doi.org/10.1038/nn.4504

Baker, N., Lu, H., Erlikhman, G., & Kellman, P. J. (2018). Deep convolutional networks do not classify based on global object shape. *PLOS Computational Biology*, *14*(12), e1006613. https://doi.org/10.1371/journal.pcbi.1006613

Bakhtiari, S., Mineault, P. J., Lillicrap, T. P., Pack, C. C., & Richards, B. A. (2021). The functional specialization of visual cortex emerges from training parallel pathways with self-supervised predictive learning. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/2021.06.18.448989

Balaguer, M. (1998, July). Philosophy of mathematics | Logic, Axioms & Proofs. https://www.britannica.com/science/philosophy-of-mathematics

Balestrieri, A., Gazzola, A., Pellitteri-Rosa, D., & Vallortígara, G. (2019). Discrimination of group numerousness under predation risk in anuran tadpoles. *Animal Cognition*, *22*(2), 223–230. https://doi.org/10.1007/s10071-019-01238-5

Bankson, B., Janini, D., & Baker, C. I. (2019). Whole-brain MEG decoding of symbolic and non-symbolic number stimuli reveals primarily format-dependent representations. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/731687

Barth, H. (2008). Judgments of discrete and continuous quantity: An illusory Stroop effect. *Cognition*, *109*(2), 251–266. https://doi.org/10.1016/j.cognition.2008.09.002

Barth, H., Kanwisher, N., & Spelke, E. S. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221. https://doi.org/10.1016/s0010-0277(02)00178-6

Barth, H., La Mont, K., Lipton, J. S., Dehaene, S., Kanwisher, N., & Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, *98*(3), 199–222. https://doi.org/10.1016/j.cognition.2004.09.011

Barth, H., La Mont, K., Lipton, J. S., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 14116–14121. https://doi.org/10.1073/pnas.0505512102

Behrens, T. E. J., Muller, T. H., Whittington, J. C. R., Mark, S., Baram, A., Stachenfeld, K. L., & Kurth-Nelson, Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron*, *100*(2), 490–509. https://doi.org/10.1016/j.neuron.2018.10.002

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Beran, M. J. (2004). Chimpanzees (Pan troglodytes) Respond to Nonvisible Sets After One-by-One Addition and Removal of Items. *Journal of Comparative Psychology*, *118*(1), 25–36. https://doi.org/10.1037/0735-7036.118.1.25

Beran, M. J. (2012). Quantity judgments of auditory and visual stimuli by chimpanzees (Pan troglodytes). *Journal of experimental psychology*, *38*(1), 23–29. https://doi.org/10.1037/a0024965

Beran, M. J., Decker, S. L., Schwartz, A. D., & Schultz, N. B. (2011). Monkeys (Macaca mulatta and Cebus apella) and human adults and children (Homo sapiens) compare subsets of moving stimuli based on numerosity. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00061

Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Bisley, J. W., & Goldberg, M. E. (2003). Neuronal activity in the lateral intraparietal area and spatial attention. *Science*, *299*(5603), 81–86. https://doi.org/10.1126/science.1077395

Bisley, J. W., & Goldberg, M. E. (2010). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, *33*(1), 1–21. https://doi.org/10.1146/annurev-neuro-060909-152823

Bracci, S., & De Beeck, H. O. (2023). Understanding human object vision: A picture is worth a thousand representations. *Annual Review of Psychology*, *74*(1), 113–135. https://doi.org/10.1146/annurev-psych-032720-041031

Brainard, D. H. (1997). The Psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897x00357

Bulthé, J., De Smedt, B., & De Beeck, H. O. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *NeuroImage*, *87*, 311–322. https://doi.org/10.1016/j.neuroimage.2013.10.049

Bulthé, J., De Smedt, B., & De Beeck, H. O. (2015). Visual number beats Abstract Numerical Magnitude: format-dependent representation of Arabic digits and dot patterns in human parietal cortex. *Journal of Cognitive Neuroscience*, *27*(7), 1376–1387. https://doi.org/10.1162/jocn_a_00787

Bulthé, J., Prinsen, J., Vanderauwera, J., Duyck, S., Daniels, N., Gillebert, C. R., Mantini, D., De Beeck, H. O., & De Smedt, B. (2019). Multi-method brain imaging reveals impaired representations of number as well as altered connectivity in adults with dyscalculia. *NeuroImage*, *190*, 289–302. https://doi.org/10.1016/j.neuroimage.2018.06.012

Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, *18*(6), 425–428. https://doi.org/10.1016/j.cub.2008.02.052

Buzsáki, G., & Moser, E. I. (2013). Memory, navigation and theta rhythm in the hippocampal-entorhinal system. *Nature Neuroscience*, *16*(2), 130–138. https://doi.org/10.1038/nn.3304

Cai, Y., Hofstetter, S., & Dumoulin, S. O. (2023). Nonsymbolic numerosity maps at the occipitotemporal cortex respond to symbolic numbers. *The Journal of Neuroscience*, *43*(16), 2950–2959. https://doi.org/10.1523/jneurosci.0687-22.2023

Cai, Y., Hofstetter, S., Harvey, B. M., & Dumoulin, S. O. (2022). Attention drives human numerosity-selective responses. *Cell Reports*, *39*(13), 111005. https://doi.org/10.1016/j.celrep.2022.111005

Cai, Y., Hofstetter, S., Van Dijk, J. A., Zuiderbaan, W., Van Der Zwaag, W., Harvey, B. M., & Dumoulin, S. O. (2021). Topographic numerosity maps cover subitizing and estimation ranges. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23785-7

Cant, J. S., & Xu, Y. (2012). Object ensemble processing in human Anterior-Medial ventral visual cortex. *The Journal of Neuroscience*, *32*(22), 7685–7700. https://doi.org/10.1523/jneurosci.3325-11.2012

Cantlon, J. F. (2017). How evolution constrains human numerical concepts. *Child Development Perspectives*, *12*(1), 65–71. https://doi.org/10.1111/cdep.12264

Cantlon, J. F., Brannon, E. M., Carter, E., & Pelphrey, K. A. (2006). Functional imaging of numerical processing in adults and 4-y-Old children. *PLOS Biology*, *4*(5), e125. https://doi.org/10.1371/journal.pbio.0040125

Cantrell, L., & Smith, L. B. (2013). Open questions and a proposal: A critical review of the evidence on infant numerical abilities. *Cognition*, *128*(3), 331–352. https://doi.org/10.1016/j.cognition.2013.04.008

Carazo, P., Font, E., Forteza-Behrendt, E., & Desfilis, E. (2009). Quantity discrimination in Tenebrio molitor: evidence of numerosity discrimination in an invertebrate? *Animal Cognition*, *12*(3), 463–470. https://doi.org/10.1007/s10071-008-0207-7

Carey, S., & Barner, D. (2019). Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences*, *23*(10), 823–835. https://doi.org/10.1016/j.tics.2019.07.004

Castaldi, E., Mirassou, A., Dehaene, S., Piazza, M., & Eger, E. (2018). Asymmetrical interference between number and item size perception provides evidence for a domain specific impairment in dyscalculia. *PLOS ONE*, *13*(12), e0209256. https://doi.org/10.1371/journal.pone.0209256

Castaldi, E., Piazza, M., Dehaene, S., Vignaud, A., & Eger, E. (2019). Attentional amplification of neural codes for number independent of other quantities along the dorsal visual stream. *eLife*, *8*. https://doi.org/10.7554/elife.45160

Castaldi, E., Piazza, M., & Eger, E. (2021). Resources underlying Visuo-Spatial working memory enable veridical large numerosity perception. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.751098

Castaldi, E., Vignaud, A., & Eger, E. (2020). Mapping subcomponents of numerical cognition in relation to functional and anatomical landmarks of human parietal cortex. *NeuroImage*, *221*, 117210. https://doi.org/10.1016/j.neuroimage.2020.117210

Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(12), 4693–4698. https://doi.org/10.1073/pnas.0600444103

Chang, C.-J., & Lin, C.-J. (2011). LIBSVM. *ACM Transactions on Intelligent Systems and Technology*, *2*(3), 1–27. https://doi.org/10.1145/1961189.1961199

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: A meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Cheng, M.-M., Mitra, N. J., Huang, X., Torr, P. H. S., & Hu, S.-M. (2015). Global Contrast based Salient Region Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *37*(3), 569–582. https://doi.org/10.1109/tpami.2014.2345401

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/s0042-6989(02)00596-5

Chung, M. K., Robbins, S., Dalton, K. M., Davidson, R. J., Alexander, A. L., & Evans, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage*, *25*(4), 1256–1265. https://doi.org/10.1016/j.neuroimage.2004.12.052

Cichy, R. M., Khosla, A., Pantazis, D., Torralba, A., & Oliva, A. (2016). Comparison of deep neural networks to spatio-temporal cortical dynamics of human visual object recognition reveals hierarchical correspondence. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep27755

Cichy, R. M., & Oliva, A. (2020). A M/EEG-FMRI fusion primer: resolving human brain responses in space and time. *Neuron*, *107*(5), 772–781. https://doi.org/10.1016/j.neuron.2020.07.001

Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. https://doi.org/10.1038/nn.3635

Cichy, R. M., Pantazis, D., & Oliva, A. (2016). Similarity-Based fusion of MEG and FMRI reveals Spatio-Temporal dynamics in human cortex during visual object recognition. *Cerebral Cortex*, *26*(8), 3563–3579. https://doi.org/10.1093/cercor/bhw135

Conrad, B., Pollack, C., Yeo, D. J., & Price, G. R. (2022). Structural and functional connectivity of the inferior temporal numeral area. *Cerebral Cortex*, *33*(10), 6152–6170. https://doi.org/10.1093/cercor/bhac492

Coubart, A., Izard, V., Spelke, E. S., Marie, J., & Stréri, A. (2013). Dissociation between small and large numerosities in newborn infants. *Developmental Science*, *17*(1), 11–22. https://doi.org/10.1111/desc.12108

Cox, D. D., & Dean, T. (2014). Neural networks and Neuroscience-Inspired computer vision. *Current Biology*, *24*(18), R921–R929. https://doi.org/10.1016/j.cub.2014.08.026

Cox, R. W. (1996). AFNI: Software for analysis and Visualization of Functional Magnetic Resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Dacke, M., & Srinivasan, M. V. (2008). Evidence for counting in insects. *Animal Cognition*, *11*(4), 683–689. https://doi.org/10.1007/s10071-008-0159-y

Dadda, M., Piffer, L., Agrillo, C., & Bisazza, A. (2009). Spontaneous number representation in mosquitofish. *Cognition*, *112*(2), 343–348. https://doi.org/10.1016/j.cognition.2009.05.009

Daitch, A. L., Foster, B. L., Schrouff, J., Rangarajan, V., Kaşikçi, I., Gattas, S., & Parvizi, J. (2016). Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(46). https://doi.org/10.1073/pnas.1608434113

Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., & Morgan, M. (2011). A common visual metric for approximate number and density. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(49), 19552–19557. https://doi.org/10.1073/pnas.1113195108

Dantzig, T., & Mazur, J. (1967). *Number: The Language of Science*. MacMillan.

Davis, H., & Albert, M. (1986). Numerical discrimination by rats using sequential auditory stimuli. *Animal Learning & Behavior*, *14*(1), 57–59. https://doi.org/10.3758/bf03200037

De Beeck, H. O. (2010). Against hyperacuity in brain reading: Spatial smoothing does not hurt multivariate fMRI analyses? *NeuroImage*, *49*(3), 1943–1948. https://doi.org/10.1016/j.neuroimage.2009.02.047

De Smedt, B., Noël, M.-P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends in Cognitive Sciences*, *7*(4), 145–147. https://doi.org/10.1016/s1364-6613(03)00055-x

Dehaene, S. (2011, April). *The number sense*. OUP USA.

Dehaene, S., & Changeux, J.-P. (1993). Development of Elementary Numerical abilities: a neuronal model. *Journal of Cognitive Neuroscience*, *5*(4), 390–407. https://doi.org/10.1162/jocn.1993.5.4.390

Dehaene, S., & Cohen, L. D. (2007). Cultural recycling of cortical maps. *Neuron*, *56*(2), 384–398. https://doi.org/10.1016/j.neuron.2007.10.004

Dehaene-Lambertz, G., Monzalvo, K., & Dehaene, S. (2018). The emergence of the visual word form: Longitudinal evolution of category-specific ventral visual areas during reading acquisition. *PLOS Biology*, *16*(3), e2004103. https://doi.org/10.1371/journal.pbio.2004103

Dembski, C., Koch, C., & Pitts, M. (2021). Perceptual awareness negativity: a physiological correlate of sensory consciousness. *Trends in Cognitive Sciences*, *25*(8), 660–670. https://doi.org/10.1016/j.tics.2021.05.009

Deng, J., Dong, W., Socher, R., Li, L., Li, K., & Li, F. (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2009.5206848

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition*, *142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

DeWind, N. K., Park, J., Woldorff, M. G., & Brannon, E. M. (2019). Numerical encoding in early visual cortex. *Cortex*, *114*, 76–89. https://doi.org/10.1016/j.cortex.2018.03.027

Ditz, H. M., & Nieder, A. (2016). Numerosity representations in crows obey the Weber–Fechner law. *Proceedings of The Royal Society B: Biological Sciences*, *283*(1827), 20160083. https://doi.org/10.1098/rspb.2016.0083

Dobs, K., Yuan, J., Martinez, J., & Kanwisher, N. (2023). Behavioral signatures of face perception emerge in deep neural networks optimized for face recognition. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(32). https://doi.org/10.1073/pnas.2220642120

Dumoulin, S. O., & Wandell, B. A. (2008). Population receptive field estimates in human visual cortex. *NeuroImage*, *39*(2), 647–660. https://doi.org/10.1016/j.neuroimage.2007.09.034

Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Current Biology*, *18*(18), R855–R856. https://doi.org/10.1016/j.cub.2008.07.053

Eger, E. (2016, January). *Neuronal foundations of human numerical representations*. https://doi.org/10.1016/bs.pbr.2016.04.015

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., & Kleinschmidt, A. (2009). Deciphering Cortical Number Coding from Human Brain Activity Patterns. *Current Biology*, *19*(19), 1608–1615. https://doi.org/10.1016/j.cub.2009.08.047

Eger, E., Pinel, P., Dehaene, S., & Kleinschmidt, A. (2013). Spatially invariant coding of numerical information in functionally defined subregions of human

parietal cortex. *Cerebral Cortex*, *25*(5), 1319–1329. https://doi.org/10.1093/cercor/bht323

Ester, E. F., Sutterer, D., Serences, J. T., & Awh, E. (2016). Feature-Selective attentional modulations in human frontoparietal cortex. *The Journal of Neuroscience*, *36*(31), 8188–8199. https://doi.org/10.1523/jneurosci.3935-15.2016

Esterman, M., & Yantis, S. (2009). Perceptual expectation evokes Category-Selective cortical activity. *Cerebral Cortex*, *20*(5), 1245–1253. https://doi.org/10.1093/cercor/bhp188

Everingham, M., Van Gool, L., Williams, C., Winn, J., & Zisserman, A. (2009). The Pascal Visual Object Classes (VOC) challenge. *International Journal of Computer Vision*, *88*(2), 303–338. https://doi.org/10.1007/s11263-009-0275-4

Faye, A., Jacquin-Courtois, S., Reynaud, E., Lesourd, M., Besnard, J., & Osiurak, F. (2019). Numerical cognition: A meta-analysis of neuroimaging, transcranial magnetic stimulation and brain-damaged patients studies. *NeuroImage: Clinical*, *24*, 102053. https://doi.org/10.1016/j.nicl.2019.102053

Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for FMRI investigations of language: defining ROIS functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194. https://doi.org/10.1152/jn.00032.2010

Ferrigno, S., Jara-Ettinger, J., Piantadosi, S. T., & Cantlon, J. F. (2017). Universal and uniquely human factors in spontaneous number perception. *Nature Communications*, *8*(1). https://doi.org/10.1038/ncomms13968

Fischl, B. (2012). FreeSurfer. *NeuroImage*, *62*(2), 774–781. https://doi.org/10.1016/j.neuroimage.2012.01.021

Fornaciai, M., Brannon, E. M., Woldorff, M. G., & Park, J. (2017). Numerosity processing in early visual cortex. *NeuroImage*, *157*, 429–438. https://doi.org/10.1016/j.neuroimage.2017.05.069

Fornaciai, M., & Park, J. (2018). Early numerosity encoding in visual cortex is not sufficient for the representation of numerical magnitude. *Journal of Cognitive Neuroscience*, *30*(12), 1788–1802. https://doi.org/10.1162/jocn_a_01320

Fornaciai, M., & Park, J. (2021). Disentangling feedforward versus feedback processing in numerosity representation. *Cortex*, *135*, 255–267. https://doi.org/10.1016/j.cortex.2020.11.013

Franconeri, S., Bemis, D. K., & Alvarez, G. A. (2009). Number estimation relies on a set of segmented objects. *Cognition*, *113*(1), 1–13. https://doi.org/10.1016/j.cognition.2009.07.002

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021, March). *Variability and consistency in early language learning*. MIT Press.

Freedman, D. J., & Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, *443*(7107), 85–88. https://doi.org/10.1038/nature05078

Fries, P. (2009). Neuronal Gamma-Band synchronization as a fundamental process in cortical computation. *Annual Review of Neuroscience*, *32*(1), 209–224. https://doi.org/10.1146/annurev.neuro.051508.135603

Fu, W., Dolfi, S., Decarli, G., Spironelli, C., & Zorzi, M. (2022). Electrophysiological signatures of numerosity encoding in a delayed Match-to-Sample task. *Frontiers in Human Neuroscience*, *15*. https://doi.org/10.3389/fnhum.2021.750582

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*(1-2), 43–74. https://doi.org/10.1016/0010-0277(92)90050-r

Gatys, L. A., Ecker, A. S., & Bethge, M. (2017). Texture and art with deep neural networks. *Current Opinion in Neurobiology*, *46*, 178–186. https://doi.org/10.1016/j.conb.2017.08.019

Gauthier, I., & Tarr, M. J. (2016). Visual object recognition: Do we (Finally) know more now than we did? *Annual review of vision science*, *2*(1), 377–396. https://doi.org/10.1146/annurev-vision-111815-114621

Gebuis, T., Gevers, W., & Kadosh, R. C. (2014). Topographic representation of high-level cognition: numerosity or sensory processing? *Trends in Cognitive Sciences*, *18*(1), 1–3. https://doi.org/10.1016/j.tics.2013.10.002

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General*, *141*(4), 642–648. https://doi.org/10.1037/a0026218

Gebuis, T., & Reynvoet, B. (2013). The neural mechanisms underlying passive and active processing of numerosity. *NeuroImage*, *70*, 301–307. https://doi.org/10.1016/j.neuroimage.2012.12.048

Geirhos, R., Rubisch, P., Michaelis, C., Bethge, M., Wichmann, F. A., & Brendel, W. (2018). ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness. *arXiv (Cornell University)*. https://arxiv.org/pdf/1811.12231.pdf

Gennari, G., Dehaene, S., Valera, C., & Dehaene-Lambertz, G. (2023). Spontaneous supra-modal encoding of number in the infant brain. *Current Biology*, *33*(10), 1906–1915.e6. https://doi.org/10.1016/j.cub.2023.03.062

Giurfa, M. (2019). An insect's sense of number. *Trends in Cognitive Sciences*, *23*(9), 720–722. https://doi.org/10.1016/j.tics.2019.06.010

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Gordon, P. (2004). Numerical Cognition Without Words: Evidence from Amazonia. *Science*, *306*(5695), 496–499. https://doi.org/10.1126/science.1094492

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Grootswagers, T., Wardle, S. G., & Carlson, T. A. (2017). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068

Gross, H., Pahl, M., Si, A., Zhu, H., Tautz, J., & Zhang, S. W. (2009). Number-Based visual generalisation in the honeybee. *PLOS ONE*, *4*(1), e4263. https://doi.org/10.1371/journal.pone.0004263

Gross, J., Price, D., & Hudson, C. (2008). *The Long Term Costs of Numeracy Difficulties* (tech. rep.). https://www.nationalnumeracy.org.uk/research-and-resources/long-term-costs-numeracy-difficulties-2008

Grotheer, M., Jeska, B., & Grill-Spector, K. (2018). A preference for mathematical processing outweighs the selectivity for Arabic numbers in the inferior temporal gyrus. *NeuroImage*, *175*, 188–200. https://doi.org/10.1016/j.neuroimage.2018.03.064

Güçlü, U., & Van Gerven, M. A. J. (2017). Increasingly complex representations of natural movies across the dorsal stream are shared between subjects. *NeuroImage*, *145*, 329–336. https://doi.org/10.1016/j.neuroimage.2015.12.036

Guggenmos, M., Sterzer, P., & Cichy, R. M. (2018). Multivariate pattern analysis for MEG: A comparison of dissimilarity measures. *NeuroImage*, *173*, 434–447. https://doi.org/10.1016/j.neuroimage.2018.02.044

Guillaumé, M., Mejias, S., Rossion, B., Dzhelyova, M., & Schiltz, C. (2018). A rapid, objective and implicit measure of visual quantity discrimination. *Neuropsychologia*, *111*, 180–189. https://doi.org/10.1016/j.neuropsychologia.2018.01.044

Hager, M. C., & Helfman, G. S. (1991). Safety in numbers: shoal size choice by minnows under predatory threat. *Behavioral Ecology and Sociobiology*, *29*(4), 271–276. https://doi.org/10.1007/bf00163984

Hannagan, T., Nieder, A., Viswanathan, P., & Dehaene, S. (2018). A random-matrix theory of the number sense. *Philosophical Transactions of the Royal Society B*, *373*(1740), 20170253. https://doi.org/10.1098/rstb.2017.0253

Harvey, B. M., & Dumoulin, S. O. (2017a). A network of topographic numerosity maps in human association cortex. *Nature Human Behaviour*, *1*(2). https://doi.org/10.1038/s41562-016-0036

Harvey, B. M., & Dumoulin, S. O. (2017b). Can responses to basic non-numerical visual features explain neural numerosity responses? *NeuroImage*, *149*, 200–209. https://doi.org/10.1016/j.neuroimage.2017.02.012

Harvey, B. M., Fracasso, A., Petridou, N., & Dumoulin, S. O. (2015). Topographic representations of object size and relationships with numerosity reveal generalized quantity processing in human parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13525–13530. https://doi.org/10.1073/pnas.1515414112

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. https://doi.org/10.1126/science.1239052

He, L., Zhang, J., Zhou, T., & Chen, L. (2009). Connectedness affects dot numerosity judgment: Implications for configural processing. *Psychonomic Bulletin & Review*, *16*(3), 509–517. https://doi.org/10.3758/pbr.16.3.509

He, L., Zhou, K., Zhou, T., He, S., & Chen, L. (2015). Topology-defined units in numerosity perception. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(41). https://doi.org/10.1073/pnas.1512408112

Hebart, M. N., Bankson, B., Harel, A., Baker, C. I., & Cichy, R. M. (2018). The representational dynamics of task and object processing in humans. *eLife*, *7*. https://doi.org/10.7554/elife.32816

Horsten, L. (2023). Philosophy of Mathematics. https://plato.stanford.edu/archives/spr2019/entries/philosophy-mathematics/

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. https://doi.org/10.1038/nrn1684

Hurewitz, F., Gelman, R., & Schnitzer, B. S. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(51), 19599–19604. https://doi.org/10.1073/pnas.0609485103

Hyde, D. C., Boas, D. A., Blair, C., & Carey, S. (2010). Near-infrared spectroscopy shows right parietal specialization for number in pre-verbal infants. *NeuroImage*, *53*(2), 647–652. https://doi.org/10.1016/j.neuroimage.2010.06.030

Im, H. Y., & Halberda, J. (2012). The effects of sampling and internal noise on the representation of ensemble average size. *Attention, Perception, & Psychophysics*, *75*(2), 278–286. https://doi.org/10.3758/s13414-012-0399-4

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–203. https://doi.org/10.1038/35058500

Izard, V. R., Sann, C., Spelke, E. S., & Stréri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10382–10385. https://doi.org/10.1073/pnas.0812142106

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLOS Biology*, *6*(2), e11. https://doi.org/10.1371/journal.pbio.0060011

Jackson-Nielsen, M., Cohen, M. A., & Pitts, M. (2017). Perception of ensemble statistics requires attention. *Consciousness and Cognition*, *48*, 149–160. https://doi.org/10.1016/j.concog.2016.11.007

Jehee, J., Brady, D., & Tong, F. (2011). Attention improves encoding of Task-Relevant features in the human visual cortex. *The Journal of Neuroscience*, *31*(22), 8210–8219. https://doi.org/10.1523/jneurosci.6153-09.2011

Jordan, K., Brannon, E. M., Logothetis, N. K., & Ghazanfar, A. A. (2005). Monkeys match the number of voices they hear to the number of faces they see. *Current Biology*, *15*(11), 1034–1038. https://doi.org/10.1016/j.cub.2005.04.056

Jordan, K., MacLean, E. L., & Brannon, E. M. (2008). Monkeys match and tally quantities across senses. *Cognition*, *108*(3), 617–625. https://doi.org/10.1016/j.cognition.2008.05.006

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The neural dynamics of attentional selection in natural scenes. *The Journal of Neuroscience*, *36*(41), 10522–10528. https://doi.org/10.1523/jneurosci.1385-16.2016

Kanwisher, N., Khosla, M., & Dobs, K. (2023). Using artificial neural networks to ask 'why' questions of minds and brains. *Trends in Neurosciences*, *46*(3), 240–254. https://doi.org/10.1016/j.tins.2022.12.008

Kasper, L., Bollmann, S., Diaconescu, A. O., Hutton, C., Heinzle, J., Iglesias, S., Hauser, T. U., Sebold, M., Manjaly, Z.-M., Pruessmann, K. P., & Stephan, K. (2017). The PhysIO toolbox for modeling physiological noise in FMRI data. *Journal of Neuroscience Methods*, *276*, 56–72. https://doi.org/10.1016/j.jneumeth.2016.10.019

Kaufman, E., & Lord, M. W. (1949). The discrimination of visual number. *American Journal of Psychology*, *62*(4), 498. https://doi.org/10.2307/1418556

Kay, K. (2018). Principles for models of neural information processing. *NeuroImage*, *180*, 101–109. https://doi.org/10.1016/j.neuroimage.2017.08.016

Kersey, A. J., & Cantlon, J. F. (2016). Neural tuning to numerosity relates to perceptual tuning in 3–6-Year-Old children. *The Journal of Neuroscience*, *37*(3), 512–522. https://doi.org/10.1523/jneurosci.0065-16.2016

Kersey, A. J., Wakim, K.-M., Li, R., & Cantlon, J. F. (2019). Developing, mature, and unique functions of the child's brain in reading and mathematics. *Developmental Cognitive Neuroscience*, *39*, 100684. https://doi.org/10.1016/j.dcn.2019.100684

Khaligh-Razavi, S.-M., Cichy, R. M., Pantazis, D., & Oliva, A. (2018). Tracking the spatiotemporal neural dynamics of real-world object size and animacy in the human brain. *Journal of Cognitive Neuroscience*, *30*(11), 1559–1576. https://doi.org/10.1162/jocn_a_01290

Khaligh-Razavi, S.-M., & Kriegeskorte, N. (2014). Deep supervised, but not unsupervised, models may explain IT cortical representation. *PLOS Computational Biology*, *10*(11), e1003915. https://doi.org/10.1371/journal.pcbi.1003915

Kim, G., Jang, J., Baek, S., Song, M., & Paik, S.-B. (2021). Visual number sense in untrained deep neural networks. *Science Advances*, *7*(1). https://doi.org/10.1126/sciadv.abd6127

King, J.-R., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. https://doi.org/10.1016/j.tics.2014.01.002

Kluth, T., & Zetzsche, C. (2016). Numerosity as a topological invariant. *Journal of Vision*, *16*(3), 30. https://doi.org/10.1167/16.3.30

Knops, A., Piazza, M., Sengupta, R., Eger, E., & Melcher, D. (2014). A shared, flexible neural map architecture reflects capacity limits in both visual Short-Term memory and enumeration. *The Journal of Neuroscience*, *34*(30), 9857–9866. https://doi.org/10.1523/jneurosci.2758-13.2014

Kobylkov, D., Mayer, U., Zanon, M., & Vallortígara, G. (2022). Number neurons in the nidopallium of young domestic chicks. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(32). https://doi.org/10.1073/pnas.2201039119

Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. https://doi.org/10.1073/pnas.0600244103

Kriegeskorte, N., & Kievit, R. A. (2013). Representational geometry: integrating cognition, computation, and the brain. *Trends in Cognitive Sciences*, *17*(8), 401–412. https://doi.org/10.1016/j.tics.2013.06.007

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, *25*, 1097–1105. http://books.nips.cc/papers/files/nips25/NIPS2012_0534.pdf

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/bf02289565

Kubilius, J., Bracci, S., & De Beeck, H. O. (2016). Deep neural networks as a computational model for human shape sensitivity. *PLOS Computational Biology*, *12*(4), e1004896. https://doi.org/10.1371/journal.pcbi.1004896

Kubilius, J., Schrimpf, M., Kar, K., Hong, H., Majaj, N. J., Rajalingham, R., Issa, E. B., Bashivan, P., Prescott-Roy, J., Schmidt, K., Nayebi, A., Bear, D. M., Yamins, D., & DiCarlo, J. J. (2019). Brain-Like Object Recognition with High-Performing Shallow Recurrent ANNs. *arXiv (Cornell University)*. https://arxiv.org/pdf/1909.06161.pdf

Kutter, E. F., Bostroem, J., Elger, C. E., Mormann, F., & Nieder, A. (2018). Single neurons in the human brain encode numbers. *Neuron*, *100*(3), 753–761.e4. https://doi.org/10.1016/j.neuron.2018.08.036

Kutter, E. F., Boström, J., Elger, C. E., Nieder, A., & Mormann, F. (2022). Neuronal codes for arithmetic rule processing in the human brain. *Current Biology*, *32*(6), 1275–1284.e4. https://doi.org/10.1016/j.cub.2022.01.054

Kutter, E. F., Dehnen, G., Borger, V., Surges, R., Mormann, F., & Nieder, A. (2023). Distinct neuronal representation of small and large numbers in the human medial temporal lobe. *Nature Human Behaviour*, *7*(11), 1998–2007. https://doi.org/10.1038/s41562-023-01709-3

Lage-Castellanos, A., De Martino, F., Ghose, G. M., Gulban, O. F., & Moerel, M. (2022). Selective attention sharpens population receptive fields in human auditory cortex. *Cerebral Cortex*, *33*(9), 5395–5408. https://doi.org/10.1093/cercor/bhac427

Lamme, V. A. F., & Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends in Neurosciences*, *23*(11), 571–579. https://doi.org/10.1016/s0166-2236(00)01657-x

Lasne, G., Piazza, M., Dehaene, S., Kleinschmidt, A., & Eger, E. (2019). Discriminability of numerosity-evoked fMRI activity patterns in human intra-parietal cortex reflects behavioral numerical acuity. *Cortex*, *114*, 90–101. https://doi.org/10.1016/j.cortex.2018.03.008

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2016). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences*, *40*. https://doi.org/10.1017/s0140525x16000960

Lemer, C., Dehaene, S., Spelke, E. S., & Cohen, L. (2003). Approximate quantities and exact number words: dissociable systems. *Neuropsychologia*, *41*(14), 1942–1958. https://doi.org/10.1016/s0028-3932(03)00123-4

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common Objects in context. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1405.0312

Lindsay, G. W. (2021). Convolutional neural networks as a model of the visual system: past, present, and future. *Journal of Cognitive Neuroscience*, *33*(10), 2017–2031. https://doi.org/10.1162/jocn_a_01544

Lipton, J. S., & Spelke, E. S. (2003). Origins of Number Sense. *Psychological Science*, *14*(5), 396–401. https://doi.org/10.1111/1467-9280.01453

Loetscher, T., Bockisch, C. J., Nicholls, M. E. R., & Brugger, P. (2010). Eye position predicts what number you have in mind. *Current Biology*, *20*(6), R264–R265. https://doi.org/10.1016/j.cub.2010.01.015

Long, B., Yu, C.-P., & Konkle, T. (2018). Mid-level visual features underlie the high-level categorical organization of the ventral stream. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(38). https://doi.org/10.1073/pnas.1719616115

Lourenco, S. F., & Longo, M. R. (2010). General magnitude representation in human infants. *Psychological Science*, *21*(6), 873–881. https://doi.org/10.1177/0956797610370158

Lucero, C., Brookshire, G., Sava-Segal, C., Bottini, R., Goldin-Meadow, S., Vogel, E. K., & Casasanto, D. (2020). Unconscious number discrimination in the human visual system. *Cerebral Cortex*, *30*(11), 5821–5829. https://doi.org/10.1093/cercor/bhaa155

Lyon, B. E. (2003). Egg recognition and counting reduce costs of avian conspecific brood parasitism. *Nature*, *422*(6931), 495–499. https://doi.org/10.1038/nature01505

Lyons, I. M., Ansari, D., & Beilock, S. L. (2014). Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping*, *36*(2), 475–488. https://doi.org/10.1002/hbm.22641

MaBouDi, H., Dona, H. S. G., Gatto, E., Loukola, O., Buckley, E., Onoufriou, P. D., Skorupski, P., & Chittka, L. (2020). Bumblebees use sequential scanning of countable items in visual patterns to solve numerosity tasks. *Integrative and Comparative Biology*, *60*(4), 929–942. https://doi.org/10.1093/icb/icaa025

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. https://doi.org/10.1093/cercor/bhg097

Melcher, D., & Piazza, M. (2011). The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. *PLOS ONE*, *6*(12), e29296. https://doi.org/10.1371/journal.pone.0029296

Merten, K., & Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience*, *21*(2), 333–346. https://doi.org/10.1162/jocn.2008.21032

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, *6*, 414–417. https://doi.org/10.1016/0166-2236(83)90190-x

Mistry, P. K., Strock, A., Liu, R., Young, G. B., & Menon, V. (2023). Learning-induced reorganization of number neurons and emergence of numerical representations in a biologically inspired neural network. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-39548-5

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M. A., & Newman, S. D. (2004). Learning to Decode Cognitive States from Brain Images. *Machine Learning*, *57*(1/2), 145–175. https://doi.org/10.1023/b:mach.0000035475.85309.1b

Moeller, S., Yacoub, E., Olman, C. A., Auerbach, E. J., Strupp, J., Harel, N., & Uğurbil, K. (2010). Multiband multislice GE-EPI at 7 tesla, with 16-fold acceleration using partial parallel imaging with application to high spatial and temporal whole-brain fMRI. *Magnetic Resonance in Medicine*, *63*(5), 1144–1153. https://doi.org/10.1002/mrm.22361

Muttenthaler, L., & Hebart, M. N. (2021). THINGSVision: a Python toolbox for streamlining the extraction of activations from deep neural networks. *Frontiers in Neuroinformatics*, *15*. https://doi.org/10.3389/fninf.2021.679838

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). https://doi.org/10.1126/sciadv.aav7903

Nelli, S., Braun, L., Dumbalska, T., Saxe, A. M., & Summerfield, C. (2023). Neural knowledge assembly in humans and neural networks. *Neuron*, *111*(9), 1504–1516.e9. https://doi.org/10.1016/j.neuron.2023.02.014

Nieder, A. (2016). The neuronal code for number. *Nature Reviews Neuroscience*, *17*(6), 366–382. https://doi.org/10.1038/nrn.2016.40

Nieder, A. (2019, November). *A brain for numbers*. MIT Press.

Nieder, A. (2020). The adaptive value of numerical competence. *Trends in Ecology and Evolution*, *35*(7), 605–617. https://doi.org/10.1016/j.tree.2020.02.009

Nieder, A., Diester, I., & Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. *Science*, *313*(5792), 1431–1435. https://doi.org/10.1126/science.1130308

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. https://doi.org/10.1126/science.1072493

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude. *Neuron*, *37*(1), 149–157. https://doi.org/10.1016/s0896-6273(02)01144-3

Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(19), 7457–7462. https://doi.org/10.1073/pnas.0402239101

Nieto-Castañón, A., & Fedorenko, E. (2012). Subject-specific functional localizers increase sensitivity and functional resolution of multi-subject analyses. *NeuroImage*, *63*(3), 1646–1669. https://doi.org/10.1016/j.neuroimage.2012.06.065

Nili, H., Wingfield, C., Walther, A., Su, L., Marslen–Wilson, W. D., & Kriegeskorte, N. (2014). A toolbox for representational similarity analysis. *PLOS Computational Biology*, *10*(4), e1003553. https://doi.org/10.1371/journal.pcbi.1003553

Norcia, A. M., Appelbaum, L. G., Ales, J., Cottereau, B. R., & Rossion, B. (2015). The steady-state visual evoked potential in vision research: A review. *Journal of Vision*, *15*(6), 4. https://doi.org/10.1167/15.6.4

Nys, J., & Content, A. (2012). Judgement of discrete and continuous quantity in adults: Number counts! *Quarterly Journal of Experimental Psychology*, *65*(4), 675–690. https://doi.org/10.1080/17470218.2011.619661

Odic, D., Valle-Lisboa, J. C., Eisinger, R. S., Olivera, M. G., Maiche, A., & Halberda, J. (2016). Approximate number and approximate time discrimination each

correlate with school math abilities in young children. *Acta Psychologica*, *163*, 17–26. https://doi.org/10.1016/j.actpsy.2015.10.010

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). COSMOMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*. https://doi.org/10.3389/fninf.2016.00027

Oosterhof, N. N., Wiestler, T., Downing, P. E., & Diedrichsen, J. (2011). A comparison of volume-based and surface-based multi-voxel pattern analysis. *NeuroImage*, *56*(2), 593–600. https://doi.org/10.1016/j.neuroimage.2010.04.270

Park, J. (2018). A neural basis for the visual sense of number and its development: A steady-state visual evoked potential study in children and adults. *Developmental Cognitive Neuroscience*, *30*, 333–343. https://doi.org/10.1016/j.dcn.2017.02.011

Park, J., DeWind, N. K., Woldorff, M. G., & Brannon, E. M. (2015). Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex*, bhv017. https://doi.org/10.1093/cercor/bhv017

Park, J., & Huber, D. E. (2022). A visual sense of number emerges from divisive normalization in a simple center-surround convolutional network. *eLife*, *11*. https://doi.org/10.7554/elife.80990

Parsons, S., & Bynner, J. (2005). *Does Numeracy Matter More?* (Tech. rep.). National Research; Development Centre for Adult Literacy; Numeracy.

Paul, J., Van Ackooij, M., Cate, T. C. T., & Harvey, B. M. (2022). Numerosity tuning in human association cortices and local image contrast representations in early visual cortex. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-29030-z

Perdue, B. M., Talbot, C. F., Stone, A., & Beran, M. J. (2012). Putting the elephant back in the herd: elephant relative quantity judgments match those of other species. *Animal Cognition*, *15*(5), 955–961. https://doi.org/10.1007/s10071-012-0521-y

Piantadosi, S. T., & Cantlon, J. F. (2017). True numerical cognition in the wild. *Psychological Science*, *28*(4), 462–469. https://doi.org/10.1177/0956797616686862

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–551. https://doi.org/10.1016/j.tics.2010.09.008

Piazza, M., De Feo, V., Panzeri, S., & Dehaene, S. (2018). Learning to focus on number. *Cognition*, *181*, 35–45. https://doi.org/10.1016/j.cognition.2018.07.011

Piazza, M., & Eger, E. (2016). Neural foundations and functional specificity of number representations. *Neuropsychologia*, *83*, 257–273. https://doi.org/10.1016/j.neuropsychologia.2015.09.025

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuospatial object individuation capacity. *Cognition*, *121*(1), 147–153. https://doi.org/10.1016/j.cognition.2011.05.007

Piazza, M., Izard, V., Pinel, P., Bihan, D. L., & Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, *44*(3), 547–555. https://doi.org/10.1016/j.neuron.2004.10.014

Piazza, M., Pinel, P., Bihan, D. L., & Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. *Neuron*, *53*(2), 293–305. https://doi.org/10.1016/j.neuron.2006.11.022

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and approximate arithmetic in an amazonian indigene group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Pinheiro-Chagas, P., Daitch, A. L., Parvizi, J., & Dehaene, S. (2018). Brain mechanisms of arithmetic: a crucial role for Ventral temporal cortex. *Journal of Cognitive Neuroscience*, *30*(12), 1757–1772. https://doi.org/10.1162/jocn_a_01319

Pisa, P. E., & Agrillo, C. (2008). Quantity discrimination in felines: a preliminary investigation of the domestic cat (Felis silvestris catus). *Journal of Ethology*, *27*(2), 289–293. https://doi.org/10.1007/s10164-008-0121-0

Potrich, D., Rugani, R., Sovrano, V. A., Regolin, L., & Vallortígara, G. (2019). Use of numerical and spatial information in ordinal counting by zebrafish. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-54740-8

Potrich, D., Sovrano, V. A., Stancher, G., & Vallortígara, G. (2015). Quantity discrimination by zebrafish (Danio rerio). *Journal of Comparative Psychology*, *129*(4), 388–393. https://doi.org/10.1037/com0000012

Proklova, D., Kaiser, D., & Peelen, M. V. (2019). MEG sensor patterns reflect perceptual but not categorical similarity of animate and inanimate objects. *NeuroImage*, *193*, 167–177. https://doi.org/10.1016/j.neuroimage.2019.03.028

Puri, A., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*(6), 607–614. https://doi.org/10.1111/j.1467-9280.2008.02130.x

Ritchie, J. B., Tovar, D. A., & Carlson, T. A. (2015). Emerging object representations in the visual system predict reaction times for categorization. *PLOS Computational Biology*, *11*(6), e1004316. https://doi.org/10.1371/journal.pcbi.1004316

Roggeman, C., Fias, W., & Verguts, T. (2010). Salience maps in parietal cortex: Imaging and computational modeling. *NeuroImage*, *52*(3), 1005–1014. https://doi.org/10.1016/j.neuroimage.2010.01.060

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLOS Biology*, *5*(8), e208. https://doi.org/10.1371/journal.pbio.0050208

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2012). Representation of numerosity in posterior parietal cortex. *Frontiers in Integrative Neuroscience*, *6*. https://doi.org/10.3389/fnint.2012.00025

Roitman, J. D., & Shadlen, M. N. (2002). Response of Neurons in the Lateral Intraparietal Area during a Combined Visual Discrimination Reaction Time Task. *The Journal of Neuroscience*, *22*(21), 9475–9489. https://doi.org/10.1523/jneurosci.22-21-09475.2002

Rousselle, L., & Noël, M.-P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology*, *44*(2), 544–560. https://doi.org/10.1037/0012-1649.44.2.544

Rubinsten, O., Korem, N., Levin, N., & Furman, T. (2020). Frequency-based Dissociation of Symbolic and Nonsymbolic Numerical Processing during Numerical Comparison. *Journal of Cognitive Neuroscience*, *32*(5), 762–782. https://doi.org/10.1162/jocn_a_01550

Rugani, R., Fontanari, L., Simoni, E., Regolin, L., & Vallortígara, G. (2009). Arithmetic in newborn chicks. *Proceedings of The Royal Society B: Biological Sciences*, *276*(1666), 2451–2460. https://doi.org/10.1098/rspb.2009.0044

Rugani, R., Regolin, L., & Vallortígara, G. (2008). Discrimination of small numerosities in young chicks. *Journal of experimental psychology*, *34*(3), 388–399. https://doi.org/10.1037/0097-7403.34.3.388

Rugani, R., Vallortígara, G., Priftis, K., & Regolin, L. (2015). Number-space mapping in the newborn chick resembles humans' mental number line. *Science*, *347*(6221), 534–536. https://doi.org/10.1126/science.aaa1379

Saad, Z. S., Reynolds, R. C., Argall, B., Japee, S., & Cox, R. W. (2005). SUMA: An interface for surface-based intra- and inter-subject analysis with AFNI. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. https://doi.org/10.1109/isbi.2004.1398837

Sawamura, H., Shima, K., & Tanji, J. (2002). Numerical representation for action in the parietal cortex of the monkey. *Nature*, *415*(6874), 918–922. https://doi.org/10.1038/415918a

Saygin, Z. M., Osher, D. E., Norton, E. S., Youssoufian, D. A., Beach, S. D., Feather, J., Gaab, N., Gabrieli, J. D. E., & Kanwisher, N. (2016). Connectivity precedes

function in the development of the visual word form area. *Nature Neuroscience*, *19*(9), 1250–1255. https://doi.org/10.1038/nn.4354

Scarf, D., Hayne, H., & Colombo, M. (2011). Pigeons on Par with Primates in Numerical Competence. *Science*, *334*(6063), 1664. https://doi.org/10.1126/science.1213357

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., & De Smedt, B. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Developmental Science*, *20*(3). https://doi.org/10.1111/desc.12372

Scott, T. L., & Perrachione, T. K. (2019). Common cortical architectures for phonological working memory identified in individual brains. *NeuroImage*, *202*, 116096. https://doi.org/10.1016/j.neuroimage.2019.116096

Sengupta, R., Surampudi, B. R., & Melcher, D. (2014). A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network. *Brain Research*, *1582*, 114–124. https://doi.org/10.1016/j.brainres.2014.03.014

Sheahan, H., Luyckx, F., Nelli, S., Teupe, C., & Summerfield, C. (2021). Neural state space alignment for magnitude generalization in humans and recurrent networks. *Neuron*, *109*(7), 1214–1226.e8. https://doi.org/10.1016/j.neuron.2021.02.004

Shum, J., Hermes, D., Foster, B. L., Dastjerdi, M. C., Rangarajan, V., Winawer, J., Miller, K. J., & Parvizi, J. (2013). A brain area for visual numerals. *The Journal of Neuroscience*, *33*(16), 6709–6715. https://doi.org/10.1523/jneurosci.4558-12.2013

Skorupski, P., MaBouDi, H., Dona, H. S. G., & Chittka, L. (2018). Counting insects. *Philosophical Transactions of the Royal Society B*, *373*(1740), 20160513. https://doi.org/10.1098/rstb.2016.0513

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Spelke, E. S., & Kinzler, K. D. (2006). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Stancher, G., Rugani, R., Regolin, L., & Vallortígara, G. (2014). Numerical discrimination by frogs (Bombina orientalis). *Animal Cognition*, *18*(1), 219–229. https://doi.org/10.1007/s10071-014-0791-7

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210*(4473), 1033–1035. https://doi.org/10.1126/science.7434014

Stoianov, I., & Zorzi, M. (2012). Emergence of a 'visual number sense' in hierarchical generative models. *Nature Neuroscience*, *15*(2), 194–196. https://doi.org/10.1038/nn.2996

Stoianov, I., & Zorzi, M. (2017). Computational foundations of the visual number sense. *Behavioral and Brain Sciences*. https://doi.org/10.1017/s0140525x16002326

Storrs, K. R., Kietzmann, T. C., Walther, A., Mehrer, J., & Kriegeskorte, N. (2021). Diverse deep neural networks all predict human inferior temporal cortex Well, after training and fitting. *Journal of Cognitive Neuroscience*, 1–21. https://doi.org/10.1162/jocn_a_01755

Strauss, M., & Curtis, L. (1981). Infant perception of numerosity. *Child Development*, *52*(4), 1146. https://doi.org/10.2307/1129500

Summerfield, C., & De Lange, F. P. (2014). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, C., Luyckx, F., & Sheahan, H. (2020). Structure learning and the posterior parietal cortex. *Progress in Neurobiology*, *184*, 101717. https://doi.org/10.1016/j.pneurobio.2019.101717

Taulu, S., Kajola, M., & Simola, J. (2003). Suppression of interference and artifacts by the signal space separation method. *Brain Topography*, *16*(4), 269–275. https://doi.org/10.1023/b:brat.0000032864.93890.f9

Templeton, C. N., Greene, E., & Davis, K. (2005). Allometry of alarm calls: Black-Capped chickadees encode information about predator size. *Science*, *308*(5730), 1934–1937. https://doi.org/10.1126/science.1108841

Testolin, A., Dolfi, S., Rochus, M., & Zorzi, M. (2020). Visual sense of number vs. sense of magnitude in humans and machines. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-66838-5

Thomas, R. K., & Chase, L. (1980). Relative numerousness judgments by squirrel monkeys. *Bulletin of the psychonomic society*, *16*(2), 79–82. https://doi.org/10.3758/bf03334444

Thompson, P., & Burr, D. (2009). Visual aftereffects. *Current Biology*, *19*(1), R11–R14. https://doi.org/10.1016/j.cub.2008.10.014

Thompson, R. F., Mayers, K. S., Robertson, R. T., & Patterson, C. J. (1970). Number coding in association cortex of the cat. *Science*, *168*(3928), 271–273. https://doi.org/10.1126/science.168.3928.271

Tsouli, A., Cai, Y., Van Ackooij, M., Hofstetter, S., Harvey, B. M., Pas, S. F. T., Van Der Smagt, M. J., & Dumoulin, S. O. (2021). Adaptation to visual numerosity changes neural numerosity selectivity. *NeuroImage*, *229*, 117794. https://doi.org/10.1016/j.neuroimage.2021.117794

Tudusciuc, O., & Nieder, A. (2007). Neuronal population coding of continuous and discrete quantity in the primate posterior parietal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(36), 14513–14518. https://doi.org/10.1073/pnas.0705495104

Tudusciuc, O., & Nieder, A. (2009). Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. *Journal of Neurophysiology*, *101*(6), 2984–2994. https://doi.org/10.1152/jn.90713.2008

Uller, C., Jaeger, R. G., Guidry, G., & Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: rudiments of number in an amphibian. *Animal Cognition*, *6*(2), 105–112. https://doi.org/10.1007/s10071-003-0167-x

Vallortigara, G. (2017). *An animal's sense of number*. Routledge.

Van Opstal, F., & Verguts, T. (2013). Is there a generalized magnitude system in the brain? Behavioral, neuroimaging, and computational evidence. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00435

Van Rinsveld, A., Guillaumé, M., Kohler, P. J., Schiltz, C., Gevers, W., & Content, A. (2020). The neural signature of numerosity by separating numerical and continuous magnitude extraction in visual cortex with frequency-tagged EEG. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(11), 5726–5732. https://doi.org/10.1073/pnas.1917849117

Van Rinsveld, A., Wens, V., Guillaumé, M., Beuel, A., Gevers, W., & De Tiège, X. (2021). Automatic processing of numerosity in human neocortex evidenced by occipital and parietal neuromagnetic responses. *Cerebral cortex communications*, *2*(2). https://doi.org/10.1093/texcom/tgab028

Verguts, T., & Fias, W. (2004). Representation of number in animals and Humans: a neural model. *Journal of Cognitive Neuroscience*, *16*(9), 1493–1504. https://doi.org/10.1162/0898929042568497

Verma, B., & Sengupta, R. (2023). Emergence of behavioral phenomena and adaptation effects in human numerosity decoder using recurrent neural networks. *Scientific Reports*, *13*(1). https://doi.org/10.1038/s41598-023-44535-3

Vialatte, F., Maurice, M., Dauwels, J., & Cichocki, A. (2010). Steady-state visually evoked potentials: Focus on essential paradigms and future perspectives. *Progress in Neurobiology*, *90*(4), 418–438. https://doi.org/10.1016/j.pneurobio.2009.11.005

Viswanathan, P., & Nieder, A. (2013). Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(27), 11187–11192. https://doi.org/10.1073/pnas.1308141110

Viswanathan, P., & Nieder, A. (2020). Spatial neuronal integration supports a global representation of visual numerosity in primate association cortices. *Journal of Cognitive Neuroscience*, *32*(6), 1184–1197. https://doi.org/10.1162/jocn_a_01548

Wagener, L., Loconsole, M., Ditz, H. M., & Nieder, A. (2018). Neurons in the endbrain of numerically naive crows spontaneously encode visual numerosity. *Current Biology*, *28*(7), 1090–1094.e4. https://doi.org/10.1016/j.cub.2018.02.023

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. https://doi.org/10.1016/j.tics.2003.09.002

Wang, L., Mruczek, R. E. B., Arcaro, M., & Kästner, S. (2014). Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, *25*(10), 3911–3931. https://doi.org/10.1093/cercor/bhu277

Wang, R., Janini, D., & Konkle, T. (2022). Mid-level Feature Differences Support Early Animacy and Object Size Distinctions: Evidence from Electroencephalography Decoding. *Journal of Cognitive Neuroscience*, *34*(9), 1670–1680. https://doi.org/10.1162/jocn_a_01883

Ward, L. M. (2003). Synchronous neural oscillations and cognitive processes. *Trends in Cognitive Sciences*, *7*(12), 553–559. https://doi.org/10.1016/j.tics.2003.10.012

Warrington, E. K., & James, M. (1967). Tachistoscopic number estimation in patients with unilateral cerebral lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, *30*(5), 468–474. https://doi.org/10.1136/jnnp.30.5.468

Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The Psychophysics of Number Representation. *Psychological Science*, *10*(2), 130–137. https://doi.org/10.1111/1467-9280.00120

Whitney, D., & Leib, A. Y. (2018). Ensemble perception. *Annual Review of Psychology*, *69*(1), 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Wurm, M. F., Tagliabue, C. F., & De Azevedo Mazza, V. (2021). Decoding location-specific and location-invariant stages of numerosity processing in subitizing. *European Journal of Neuroscience*, *54*(3), 4971–4984. https://doi.org/10.1111/ejn.15352

Xiao, J., Hays, J., Ehinger, K. A., Oliva, A., & Torralba, A. (2010). SUN database: Large-scale scene recognition from abbey to zoo. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2010.5539970

Xie, S., Hoehl, S., Moeskops, M., Kayhan, E., Kliesch, C., Turtleton, B., Köster, M., & Cichy, R. M. (2022). Visual category representations in the infant brain. *Current Biology*, *32*(24), 5422–5432.e6. https://doi.org/10.1016/j.cub.2022.11.016

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. https://doi.org/10.1016/s0010-0277(99)00066-9

Yamins, D., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, *19*(3), 356–365. https://doi.org/10.1038/nn.4244

Yamins, D., Hong, H., Cadieu, C. F., Solomon, E. A., Seibert, D., & DiCarlo, J. J. (2014). Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8619–8624. https://doi.org/10.1073/pnas.1403112111

Yeo, D. J., Pollack, C., Merkley, R., Ansari, D., & Price, G. R. (2020). The "Inferior Temporal Numeral Area" distinguishes numerals from other character categories

during passive viewing: A representational similarity analysis. *NeuroImage*, *214*, 116716. https://doi.org/10.1016/j.neuroimage.2020.116716

Yeo, D. J., Wilkey, E. D., & Price, G. R. (2017). The search for the number form area: A functional neuroimaging meta-analysis. *Neuroscience & Biobehavioral Reviews*, *78*, 145–160. https://doi.org/10.1016/j.neubiorev.2017.04.027

Young, E. D. (1998). Parallel processing in the nervous system: Evidence from sensory maps. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 933–934. https://doi.org/10.1073/pnas.95.3.933

Zhang, J., Ma, S., Sameki, M., Sclaroff, S., Betke, M., Lin, Z., Shen, X., Price, B., & Měch, R. (2015). Salient Object Subitizing. *Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2015.7299031

Zhang, X., & Wu, X. (2020, November). On Numerosity of Deep Neural Networks. https://arxiv.org/abs/2011.08674

Zhou, C., Xu, W., Liu, Y., Xue, Z., Chen, R., Zhou, K., & Liu, J. (2021). Numerosity representation in a deep convolutional neural network. *Journal of Pacific Rim Psychology*, *15*, 183449092110126. https://doi.org/10.1177/18344909211012613

Zhuang, C., Yan, S., Nayebi, A., Schrimpf, M., Frank, M. C., DiCarlo, J. J., & Yamins, D. (2021). Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(3). https://doi.org/10.1073/pnas.2014196118

Zorzi, M., Testolin, A., & Stoianov, I. (2013). Modeling language and cognition with deep unsupervised learning: a tutorial overview. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00515

Zuiderbaan, W., Harvey, B. M., & Dumoulin, S. O. (2012). Modeling center-surround configurations in population receptive fields using fMRI. *Journal of Vision*, *12*(3), 10. https://doi.org/10.1167/12.3.10