Processing of Numbers by Single Neurons in the Human Medial Temporal Lobe

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von

Esther F. Kurz

aus Friedrichroda

Tübingen
2023

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 15.10.2024

Dekan: Prof. Dr. Thilo Stehle

1. Berichterstatter: Prof. Dr. Andreas Nieder

2. Berichterstatter: Prof. Dr. Steffen Hage

3. Berichterstatter: Prof. Dr. Gregor Rainer

Erklärung

Ich erkläre hiermit, dass ich die zur Promotion eingereichte Arbeit mit dem Titel:

„Processing of Numbers by Single Neurons in the Human Medial Temporal Lobe“

selbständig verfasst, nur die angegebenen Quellen und Hilfsmittel benutzt und wörtlich oder inhaltlich übernommene Stellen als solche gekennzeichnet habe. Ich erkläre, dass die Richtlinien zur Sicherung guter wissenschaftlicher Praxis der Universität Tübingen (Beschluss des Senats vom 25.05.2000) beachtet wurden. Ich versichere an Eides statt, dass diese Angaben wahr sind und dass ich nichts verschwiegen habe. Mir ist bekannt, dass die falsche Abgabe einer Versicherung an Eides statt mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft wird.

Tübingen, den 12.10.2023 .....

Esther F. Kurz

Contents

Zusammenfassung 9

Abstract 11

I. Synopsis 13

1. Introduction 15

1.1. Representational Systems for Quantity 15
1.1.1. Mental Number Line and Analogue Magnitude System (ANS) . 15
1.1.2. The 'Number Sense' as Ancient Evolutionary Heritage 18
1.1.3. Exact Representation of Quantity 21
1.2. The Core Number System 23
1.2.1. The Triple-Code Model for Numerical Cognition 23
1.2.2. Characteristics of Neurophysiological Numerical Representations 24
1.2.3. The Magnitude Code: Abstract or Not? 27
1.2.4. A More General Sense of Number 32
1.3. Subitizing in Enumeration Processes 33
1.3.1. Subitizing versus Estimation: One or Two Representational Systems? 33
1.3.2. The Role of Nonnumerical Mechanisms in Subitizing 35
1.3.3. Neuronal Correlates for Subitizing 36
1.4. Mathematics in the Brain 37
1.4.1. Core Brain Regions for Mental Arithmetic 38
1.4.2. The Medial Temporal Lobe (MTL) in Mathematical Cognition . . 39
1.4.3. Monkey Single Neurons Encoding Abstract Rules 41
1.5. Working Memory in the Medial Temporal Lobe 44
1.5.1. Different Neuronal Codes for Memories 44
1.5.2. Functional Implications 45
1.6. Motivation 46

2. Main Results 48

2.1. Single Neurons in the Human Brain Encode Numbers 48
2.1.1. Single Neurons Respond to Nonsymbolic Numerosities 48
2.1.2. Single-Cell Responses to Symbolic Numerals 49
2.1.3. Neuronal Population Coding 49
2.1.4. Number Encoding in Later Task Phases 51

2.2. Distinct Neuronal Representation of Small and Large Numbers in the Human Medial Temporal Lobe 51
2.2.1. Behaviour 51
2.2.2. Neuronal Responses and Tuning Characteristics 52
2.2.3. Coding Differences at the Population Level 53
2.3. Neuronal Codes for Arithmetic Rule Processing in the Human Brain . . 54
2.3.1. Single Neurons Respond to Calculation Rules 54
2.3.2. Notation-Independent Representation of Addition and Subtraction Rules 55
2.3.3. Cross-Notation Decoding of Addition and Subtraction 56
2.3.4. Cross-Temporal Calculation Rule Decoding 56

3. Discussion **58**
3.1. Number Neurons in the Human Brain 58
3.1.1. Encoding of Numerical Information by MTL Single Neurons . . 59
3.1.2. Segregated Populations of Numerosity- and Numeral-Selective Cells 59
3.1.3. Neuronal Codes for Numbers 60
3.1.4. Distinct Representations of Nonsymbolic and Symbolic Quantities 61
3.1.5. Coding Dichotomy for Small and Large Numerosities 62
3.2. Neuronal Codes for Abstract Arithmetic Rules 63
3.2.1. Single Neurons Selective for Numerical Rules 63
3.2.2. Online Maintenance of Rule Information in Working Memory . 64
3.2.3. Static and Dynamic Codes for Working Memory 65
3.3. Conclusion 65

Abbreviations **67**

References **68**

II. Individual Publications 85

Statement of Contributions **87**

Publication 1: Single Neurons in the Human Brain Encode Numbers **89**

Publication 2: Neuronal Codes for Arithmetic Rule Processing in the Human Brain **117**

Publication 3: Distinct Neuronal Representation of Small and Large Numbers in the Human Medial Temporal Lobe **141**

Zusammenfassung

Jahrzehnte der Forschung haben gezeigt, dass uns allen ein intuitives Zahlenverständnis innewohnt, das in unserem Gehirn innerhalb eines weit verteilten und doch eng verzahnten neuronalen 'Zahlen-Netzwerks' fest verankert ist. Erkenntnisse über dieses hochkomplexe System stammen im Wesentlichen aus zwei Quellen: Tiermodelle erlauben es, neuronale Korrelate auf der Ebene einzelner Nervenzellen zu erfassen; das menschliche Gehirn hingegen wird überwiegend mittels bildgebender Verfahren erforscht. Im Rahmen dieser Doktorarbeit konnte ich die seltene Gelegenheit ergreifen, die Aktivität von einzelnen Neuronen im Medialen Temporallappen (MTL) neurologischer Patienten zu messen, und dadurch entscheidende Vorzüge der beiden vorgenannten Ansätze vereinigen. Mittels zweier Aufgabenstellungen – die Probanden wurden gebeten entweder einfache Rechenaufgaben zu lösen oder die Parität (gerade/ungerade) einer Zahl anzugeben – konnte ich verschiedene Aspekte unserer numerischen Fähigkeiten erforschen, die den konventionellen Ansätzen bisher verborgen geblieben sind.

In einem ersten entscheidenden Schritt konnte ich zeigen, dass Einzelzellen des MTL in der Lage sind, Anzahlen und einfache arithmetische Konzepte zu kodieren. Dabei wurden zufällige Punktmuster und Arabische Ziffern von zwei separaten Zellpopulationen repräsentiert. Die Antwortprofile der beiden unterlagen einem einheitlichen 'Labelled-Line Code', gleichzeitig beobachtete ich aber auch deutliche Unterschiede, insbesondere im Hinblick auf die Präzision der Repräsentationen. Die gefundenen Zahlen- und Regelzellen könnten die fundamentalen Bausteine darstellen, über die sich uns die Türen in die komplexe und höchst abstrakte Welt der Mathematik öffnen. Im Weiteren untersuchte ich das Phänomen des 'Subitizing': Kleine Zahlen können 'auf einen Blick' erfasst werden, größere hingegen müssen langsam gezählt oder fehleranfällig geschätzt werden. Diese Dichotomie spiegelte sich auch im Antwortverhalten der Zellen wider und liefert einen entscheidenden Hinweis auf das komplexe Zusammenspiel von Aufmerksamkeit, Arbeitsgedächtnis und Zahlenrepräsentationen, dem dieses Phänomen zu Grunde zu liegen scheint. Schlussendlich beobachtete ich, dass verschiedene Bereiche des MTL die arithmetischen Regeln auf unterschiedliche Weise in ihren Zellpopulationen zwischenspeichern. Dieser Befund zeigt, dass dieses hoch assoziative Areal als wesentlicher Bestandteil des kortikalen 'Mathe-Netzwerks' insbesondere auch in Prozessen des Arbeitsgedächtnisses eine wichtige Rolle zu spielen scheint. All diese Erkenntnisse liefern wertvolle Puzzleteile, die das Wissen um unser Zahlenverständnis weiter vertiefen.

Abstract

Decades of research have shown that animals and humans alike share an innate 'sense of number' that provides the cognitive start-up tool for the construction of all formal mathematical concepts. This system is anchored in a complex, highly distributed and interconnected neuronal 'number network'. In this thesis I could bridge the gap between single-unit recordings in animals and macroscopic functional imaging studies in humans, using the rare opportunity to record the activity of single neurons in the medial temporal lobe (MTL) of behaving human patients. In two different experimental protocols, calculation task and parity judgement task, we were able to explore several seemingly disparate aspects of numerical and mathematical cognition, addressing questions that had yet been eluded from investigation.

As a first fundamental insight, we showed that single cells in the MTL can encode information about both quantities and simple arithmetic rules. These numerical representations follow a labelled-line coding. Segregated populations of neurons that encode numerosities and numerals with distinct tuning profiles, however, indicate different degrees of abstractness for nonsymbolic and symbolic stimulus formats. As a neuronal basis of numerical and arithmetic representations these cells may ultimately give rise to number theory and mathematics. Next, we revealed striking coding differences between small and large numerosities, mirroring subitizing and estimation processes that provide an intriguing link to the complex interplay of attention, working memory, and number representations. Finally, we uncovered static and dynamic coding mechanisms in different subregions of the MTL that do not only emphasize the MTL's role as an integral part of a wider cortical maths network, but equally important, highlight the substantial role this highly associative area also plays in working memory processes. All these findings provide valuable puzzle pieces that deepen our understanding of numerical representations constituting our 'sense of number'.

Part I.
Synopsis

1. Introduction

Decades of research on numerical cognition have shown that we all share an innate 'sense of number' (Dehaene, 1997) which is anchored in a highly distributed and interconnected neuronal 'number network'. In combination with our ability to flexibly adapt our behaviour according to given rules, this complex system provides the foundation for the construction of all formal mathematical concepts. So far, the search for neuronal correlates of numerical representations has spawned countless many-faceted and most diverse findings, deepening our understanding of the underlying brain mechanisms. At the same time, it has become more and more evident that 'number' is not merely an abstract concept but can only be accounted for when considering it as a unifying, overarching principle governing all cognitive domains – from perception to action, and everything in between.

1.1. Representational Systems for Quantity

Numbers are ubiquitous: They do not only open the doors to science and technology, but are also an integral part of our everyday life. We use them to label and distinguish different objects or persons (for example, the shirt numbers of football players), to tell time and date, to compare prices and discuss stock quotations, to quantify everything imaginable. What all these examples have in common is that they usually come as number words or Arabic digits. To represent and manipulate quantities in such a *symbolic* form is a uniquely human cultural achievement that takes years of education to master.

By the time children go to elementary school, however, they do not 'start at zero', but already understand the concept of *'cardinality'*, i.e. they can assess the number of countable elements in a set (also called *numerosity*), and their mathematical knowledge includes already even more complex concepts like (approximate) addition and subtraction (amongst others) (Gelman and Gallistel, 1978). This indicates that symbolic representations and fundamental arithmetic abilities build on intuitive *nonsymbolic* numerical representations that do not rely on counting or number symbols.

1.1.1. Mental Number Line and Analogue Magnitude System (ANS)

One of the most fundamental assumptions about internal numerical representations is the idea of a *'mental number line'* on which quantities are represented according to ordinal numerical distance. Indeed, humans seem to share the intuition that numbers map onto space. An extensively studied example for this relationship is the *'spatial numerical association of response codes' (SNARC) effect* (Dehaene et al., 1993), the observation that small numbers elicit faster leftward responses and large numbers faster rightward responses, even when magnitude is irrelevant. This effect seems to be

rather amodal (it was observed, e.g. for Arabic numbers, number words, dot patterns or even auditory stimuli; Nuerk et al., 2005; Zhou et al., 2016), but the directionality of mapping is strongly influenced by culture-specific experiences and the dominant context in which numerical notations are typically presented: Dehaene et al. (1993) observed their left-to-right mapping in Western students who usually write from left to right, but reported reversed directionality in right-to-left writing Iranian subjects. Hung et al. (2008) showed that Chinese students aligned Arabic numerals horizontally from left to right, but simple Chinese number words vertically from top to bottom.

That this number-space mapping is not merely a cultural invention was convincingly demonstrated by Dehaene et al. (2008) who asked members of the Amazonian indigenous Mundurucú group to perform a nonsymbolic number mapping task. For that, a line segment was presented, with one dot at the left and ten dots at the right end (left-to-right directionality was thus predetermined by the experimenters but of no concern in this context). Then, intermediate numbers were presented and the subjects were asked to point to the corresponding location on the line segment. Although most of the subjects had little formal education and no access to rulers or other measurement devices, they required only a minimal instruction period to understand that number can be mapped onto a spatial scale.

In the *'analogue number system' (ANS)* (see Nieder, 2016a; Feigenson et al., 2004, for a review), numerosities are then represented as approximate, noisy estimates along the mental number line. In other words, our brain builds mental representations like, for example, 'five-ish' that are maximally activated for the (preferred) quantity '5', but to some extent also when we perceive quantities '4' or '6', and to an even lesser extent for quantities '1' or '8'. With increasing magnitudes, these estimates become systematically less precise; when being presented with, for example, eight objects our guesses of set size may vary between 7, 8 and 9, in contrast, when seeing 30 objects it is more likely that our estimates come from a much broader range like 25 to 35.

Following this relationship, two phenomena are typically observed during numerical judgements, both in adults (Moyer and Landauer, 1967; Buckley and Gillman, 1974) and children (Sekuler and Mierkiewicz, 1977; Duncan and McFarland, 1980) as well as in nonhuman primates (Washburn and Rumbaugh, 1991; Dehaene et al., 1998; Brannon and Terrace, 1998). First, as mental representations of more distant numbers are more unequivocal, it is easier to discriminate two numerically distant numbers (e.g. 4 – 8) than two numerically closer numbers (e.g. 7 – 8), a finding called the *'numerical distance effect'*. And second, the *'numerical size effect'*, the finding that it is easier to discriminate two small numbers (e.g. 3 – 4) than two large numbers (e.g. 9 – 10) at a given numerical distance (1 in this example). As mental representations of large numbers share increasingly more overlap with other nearby numbers, the likelihood of confusion increases. Closely related to this is another critical signature of the ANS.

Especially for large quantities that prevent exact counting, successful discrimination does not depend on absolute numerosities but rather on the ratio by which two set sizes differ: Although containing much more elements, comparing 100 – 200 is as difficult as 50 – 100 (both instantiate a 1:2 ratio). The finest numerical ratio that can still be discriminated is a subjective sensation and serves as a measure for the precision of an individual's number system.

Although numerical judgements are clearly different from sensory processes – when assessing numerosity, the concrete physical appearance of the stimulus is meaningless –, number perception shares many similarities with lower-level sensory representations; the magnitude effects described above are but one example (see, however, also section 1.2.4). Like other physical magnitudes, number perception follows the classical psychophysical *Weber-Fechner law* (Fechner, 1860) which postulates that a subjective sensation S is a logarithmic function, $S = k * \log(I)$, of objective stimulus intensity I. Accordingly, the minimum amount of change ΔI between two magnitudes I and $I + \Delta I$ that has to occur for reliable discrimination is proportional to I (*Weber's law*), resulting in the constant *Weber fraction* $w = \Delta I / I$ (Weber, 1850). Indeed, this 'just noticeable difference' is directly linked to the ratio-dependence discussed above, as the Weber fraction is derived from the finest numerical ratio that can still be discriminated.

The analogue magnitude system can be described by different mathematical models. The *'scalar-variability' model* assumes that the number line is linearly scaled and each numerosity is described by a Gaussian 'tuning function' whose variability scales linearly with numerosity (Meck and Church, 1983; Gallistel and Gelman, 1992). The alternative *'log-Gaussian' model* assumes a logarithmically compressed scaling with Gaussian tuning functions of constant variability (corresponding to the Weber fraction) (Nieder and Miller, 2003; Dehaene, 2007; Merten and Nieder, 2009). Indeed, the behavioural predictions are essentially identical for both models, differing only subtly in terms of asymmetries in the tuning curves. A (logarithmically) compressed scaling of stimulus magnitudes following Weber's law, however, provides considerable advantages in terms of neuronal processing as it enlarges the coding space to deal also with stimuli of unlimited range (like numerosities), and is also commonly found throughout the sensory system (Dayan and Abbott, 2005). (Although the log-Gaussian model does not predict an upper limit for number representations, Anobile et al., 2016, argued that the perception of textures, i.e. stimuli containing so many and densely packed items that they cannot be individuated anymore, poses a distinct representational regime that follows different dependencies and psychophysical laws.)

Finally, the analogue magnitude system is characterized by another important feature: modality-independence. 'Numerosity' is a highly abstract quantitative concept

that can come in many different forms: four objects of whatever shape, simultaneously presented and enumerable at a single glance, or sequentially presented across time; four light flashes, tones, brush strokes or taps with a finger. Additionally, we humans have developed different symbolic notations, most dominantly digits like '4' or 'IV' (restricted to the visual modality), and number words in different languages like 'four', 'vier' or 'quatre' (usually encountered orally, and less frequently visually in written form) (see also section 1.1.3). All these stimuli share the cardinality 'four', irrespective of the sensory appearance of the elements.

Interestingly, numerical distance/size effects as well as ratio-dependence are observed in countless studies using different kinds of stimuli. For example, when asked to indicate the numerically larger of two stimuli, Buckley and Gillman (1974) observed the numerical distance effect for symbolic Arabic numerals and nonsymbolic dot patterns of various configurations alike (though less pronounced for the numerals). Whalen et al. (1999), in contrast, assessed numerosity via hand movements (asking their subjects to press a key a specific number of times as fast as possible), or by watching sequences of rapidly flashing lights, which required integration over time to approximate numerosity. Arrighi et al. (2014), finally, showed that adaptation to number (reflected in an underestimation of set sizes after adaptation to a high numerosity) occurred across sensory modalities (auditory/visual) and presentation formats (sequential/simultaneous).

All these studies argue for a generalized sense of number that integrates quantity information across space, time and even modality. It must be noted, however, that behavioural similarities do not necessarily imply a unitary and/or exclusively abstract *neuronal* representation for all this different forms of representations (as discussed in section 1.2.3), but rather highlight the origins of the nonsymbolic capabilities inherited from our animal ancestors (see section 1.1.2).

1.1.2. The 'Number Sense' as Ancient Evolutionary Heritage

There is plenty of accumulating evidence that the ANS serves as a critical ontogenetical building block for our later abstract mathematical abilities that is qualitatively transformed and 'harnessed' when children learn mathematics in school (Piazza, 2010).

Gilmore et al. (2010), for example, showed that pre-school children succeeded in approximately adding and comparing nonsymbolic numerical quantities, although still lacking any relevant mathematical education. Performance, however, strongly depended on the ratio of the two sets sizes. Most importantly, individual performance correlated with achievements in later conducted mathematics tests and the children's mastery of number words and symbols, independent of achievements in reading or general intelligence. In line with this findings, Halberda et al. (2008) reported that

nonsymbolic estimation skills measured in 9th grade retrospectively correlated with the children's past scores in symbolic maths tests assessed in a longitudinal study, in every year of testing, extending all the way back to kindergarten. Starr et al. (2013) went even further back and showed that even in 3.5-year-olds preverbal number sense and mathematical abilities are positively related.

Indeed, even prelinguistic infants can discriminate the numerosity of different objects. Starkey et al. (1990), for example, showed 6-month-old infants multiple displays of two or three heterogeneous household objects. During this habituation phase, the infants' looking times at each display gradually decreased over time. Surprisingly, when presenting alternating displays of either two or three objects during the subsequent test phase, infants looked longer at the displays with the novel number of objects. The authors concluded, that infants successfully responded to set size irrespective of other nonnumerical but perceptually salient differences.

Importantly, infants' number discrimination abilities are not limited to small numerosities (Feigenson et al., 2004; McCrink and Wynn, 2004). Using the same habituation approach but with more rigorous controls for perceptual variables, further experiments showed that infants can discriminate even large sets of up to 32 objects. The success, however, depended strongly on age and ratio of the set sizes: While 6-month-olds successfully discriminate numerosities with a 1:2 ratio (Xu and Spelke, 2000; Xu, 2003), only 10-month-olds succeed also with a 2:3 ratio (Xu and Arriaga, 2007). With increasing age, the 'resolution' of the ANS further sharpens such that adults can reliably discriminate numerosity ratios as fine as 9:10 (van Oeffelen and Vos, 1982; Halberda and Feigenson, 2008).

Wynn (1992), finally, showed that infants do not only understand basic numerical relationships, but are also able to manipulate these concepts in numerically meaningful ways. When 5-month-old infants observed the mathematical operation '$1 + 1$' or '$2 - 1$', playfully performed in a child-appropriate setting, they looked longer at incorrect outcomes than at the correct one. These findings indicate that infants do not simply expect an arithmetical operation to result in a numerical change, but that they form expectations about both size and direction of change.

All these findings corroborate the idea that this intuitive understanding of a 'number concept' is innate, instead of merely emerging through individual learning in the context of cultural transmission and education.

Further evidence that the analogue magnitude system is deeply rooted in our ancestry comes from two seminal studies that showed that even indigenous people lacking formal mathematical education and a verbal counting system are able to process large numbers far beyond their naming range.

Gordon (2004), for example, asked members of the Amazonian Pirahã tribe to repli-

cate numerosities 2 to 10 by placing tokens in one-to-one correspondence with tokens of the to-be-counted group. Accuracy was relatively good with up to 2 or 3 tokens, but then deteriorated considerably with increasing numerosity, showing that this apparently simple matching task posed a great challenge for the participants. Interestingly, their attempts to get the answers correct were not random: The standard deviation of their estimates increased proportionally with set size, resulting in a constant discrimination ratio of approximately 6:7.

These results are comparable to Pica et al. (2004) who reported the classical numerical distance effect and a similar amount of imprecision in performance for Amazonian Mundurucú people who had to compare quantities of 20 to 80 dots (thus, far beyond anybody's counting range) in simple number comparison and also more complex approximate addition tasks. Interestingly, Whalen et al. (1999) who asked Western college students to approximate numerosities either by producing target numbers of key presses as fast as possible, or by watching sequences of rapidly flashing lights – thus, assessing ANS precision in a less familiar way than the 'classical' dot array displays – observed a discrimination ratio comparable to the indigenous people.

This indicates that ANS precision is strongly shaped by practice and increased engagement in numerical discrimination during education of school mathematics (Booth and Siegler, 2006; Halberda and Feigenson, 2008).

Estimation skills, finally, are ubiquitously found throughout the animal kingdom and have been reported not only for mammals like primates, dolphins, lions and rats, but also birds like pigeons and corvids, and even fish, amphibians, molluscs and insects (Nieder, 2021). Clearly, the intrinsic understanding of numerosity poses an evolutionary advantage as it allows for more informed decisions, for example, when choosing more profitable food zones or when deciding whether to attack or retreat from a group of competitors.

Especially in primates, number estimation has been studied extensively over the last decades in strictly controlled experimental setups, revealing impressive numerical skills. Chimpanzee Sarah, for example, successfully matched exemplars of different proportions and quantities, even when sample stimulus and the alternatives she could choose from were physically highly dissimilar (e.g. food items, wooden disks or liquid-filled jars), indicating the presence of a basic concept of numerosity and proportion (Woodruff and Premack, 1981). Going one step further, Rumbaugh et al. (1987) let chimpanzees choose between two trays, each of which contained two wells with variable quantities of food items. The animals reliably chose the tray containing the most food items, showing that the animals successfully 'summed up' the contents of the food wells on each tray. These results may, to some extent, be explainable by the extensive training these animals experienced in all kinds of cognitive tasks (Woodruff and Premack, 1981, for example, reported that four other chimpanzees failed in their

match-to-quantity task). But still, they show that primates possess the cognitive prerequisites for such advanced numerical tasks.

Brannon and Terrace (1998), finally, showed that rhesus monkeys can spontaneously represent also *abstract* numerosities and apply simple ordinal rules to them. In their seminal work, the authors first trained the animals to order stimuli of the numerosities 1 to 4; abundant controls for nonnumerical features of the stimuli (like size, equal surface area, shape and colour) ensured that behaviour was exclusively controlled by quantity. In a second task, the monkeys were then required to respond in an ascending numerical order to stimulus pairs of numerosities ranging from 1 to 9. Interestingly, the animals succeeded not only for stimuli from the familiar number range, but also for stimulus pairs that contained novel quantities from the extended range; an effect that cannot be attributed to prior training. Importantly, accuracy reflected the afore-mentioned numerical distance effect. These results indicate that the monkeys proficiently detected ordinal disparities also among novel numerosities, and that the ordinal rule learned in training was then readily extrapolated to the extended number range. Finally, when directly comparing monkeys' and humans' performance in this ordinal comparison task (using an even broader number range), striking qualitative and quantitative similarities were observed (Cantlon and Brannon, 2006).

These impressive capabilities for numerical processing indicate not only that already our animal ancestors possessed a number sense, but also that we preserved this cognitive tool in our evolution to culturally and scientifically advanced humans.

1.1.3. Exact Representation of Quantity

The approximate number system may be an excellent cognitive tool whenever quick, imprecise estimates of quantities suffice to arrive at informed decisions: In the race for the bowl with the most candies, it is irrelevant whether you fight for 19 or 20 pieces – all that matters is that you are faster than your siblings so as not to end up with the bowl containing only 3 or 4 pieces. In addition to this fast and error-prone estimation process, we humans have found a way to assess also the *exact* cardinality of a set of objects via slow *counting* routines (Gelman and Gallistel, 1978).

Although the focus of this thesis is not on counting, it should be emphasized that this process poses a fundamentally different approach to assess set size. Studies with indigenous people (Gordon, 2004; Pica et al., 2004) and young children (Wynn, 1990) show that – unlike estimation (see section 1.1.2) – counting is nothing intuitive, but needs to be learned progressively and effortfully (Carey, 2009; Carey and Barner, 2019). Furthermore, counting requires attention for the active individuation of all elements in a set, and, subsequently, working memory capacities (especially subvocalization components) (Logie and Baddeley, 1987; Leybaert and van Cutsem, 2002) to keep a running total while successively integrating them (Piazza and Izard, 2009).

Finally, counting puts the individuated items into one-to-one correspondence with members of a set of number tags – typically number words or digits (Gelman and Gallistel, 1978) – and is, thus, inextricably linked to symbolic number representations.

Indeed, it was the insight that *any* series of identical tokens (e.g. notches on wood or knots in a rope) can be used as number tags to represent a collection of objects of whatever form (e.g. fruit, people, animals) exactly and durably that laid the foundation for the evolution of symbolic numerical representations. And humans could never have walked on the moon if mankind had not developed an abstract and highly precise system of representing numbers, manipulable according to most sophisticated and advanced sets of mathematical principles.

Most modern Western societies rely on a sophisticated *symbolic enumeration system* which, upon closer inspection, poses a complex calculation in itself. The numeral '274', for example, corresponds to the number word 'two-hundred-and-seven-ty-four', and is a specific decomposition into the quantity $2 * 100 + 7 * 10 + 4 * 1$ following a positional additive-multiplicative notational syntax: Different orders of magnitudes are referred to by a fundamental 'base' number word; each base, in turn, is preceded by another number word which indicates how many times the base value is counted. In the end, all these 'base products' are added up to unequivocally obtain the quantity of the given number. Verbally, this base-10 system (in its simplest form) radically reduces the numerical lexicon to a set of basic number words for quantities 1–9 ('one' to 'nine') and some multiplier words ('ten' or suffix '-ty', 'hundred', 'thousand', etc.). For numerical symbols, this system was even further perfected by the introduction of a positional notation, i.e. each base occupies a well-defined place in the multidigit numeral (which makes it superfluous to write out the multiplier), and, importantly, the invention of the special symbol '0' serving as a placeholder that explicitly indicates the absence of a given base (Boyer, 1944, 1968).

What may seem cumbersome at first glance is actually the result of a millenium-old evolution of attempts to communicate about and keep permanent record of quantities, thereby exploiting the capabilities of our cognitive system as efficiently as possible (Dehaene, 1997). Although our long-term memory has extraordinary capacities, it is impossible to memorize a unique number word for *every* single quantity. At the same time, producing or perceiving undifferentiated series of identical tokens (probably the first form of durable 'symbolic' representations) becomes very tedious and error-prone for large magnitudes given the sharp restrictions of our visual system and short-term memory (see also section 1.3.3). Retrieving the meaning of an arbitrary shape, in contrast, is not only faster but also far more accurate. Thus, expressing large nonsymbolic quantities as combinations of a few well-defined *symbols* corresponding to a readily comprehensible range of small numerosities allows overcoming these cognitive limitations.

This excursus into the history of number shall emphasize that symbolic representations differ from nonsymbolic numerosities in two fundamental aspects: First, their physical properties, e.g. the arbitrary shapes of Arabic digits, do not intrinsically correlate with the quantity they denote. For example, that patterns of 19 and 20 dots show consecutive quantities can be derived directly from the fact that the former stimulus contains one more element than the latter one. In contrast, recognising that the completely dissimilar shapes '19' and '20' denote also consecutive numbers requires not only the comprehension of the semantic numerical meaning of the individual symbols but also a full understanding of the logic behind the additive-multiplicative enumeration system. This is an active and effortful learning process (Piazza, 2010). Second, and probably more importantly, symbolic representations are *per definition* exact – while the ANS representation of 'twenty-ish' may cover also quantities 18 to 22, the digit '20' refers exclusively to the quantity 20. These aspects should be kept in mind when trying to explore the neuronal correlates underlying symbolic representations.

1.2. The Core Number System

We as educated human adults have extensive experience with numerical information that we encounter daily not only as number words and symbolic numerals but also as nonsymbolical quantities of various forms. Over the last decades, researchers have tried to unravel how these different 'numerical codes' interact and are represented internally by our brain.

1.2.1. The Triple-Code Model for Numerical Cognition

In the early 1990s, Dehaene (1992) meticulously gathered behavioural and neuropsychological evidence from brain-lesioned, aphasic and healthy human adults as well as developing children and animals, for both exact and approximate numerical abilities, and synthesized his findings in the seminal functional *triple-code model for numerical cognition*.

This model is based on two central assumptions. First, there is no single central number representation, instead, numbers are represented in three different codes, namely, an auditory verbal word frame (e.g. /six/), a visual Arabic number form (e.g. '4'), and an analogue magnitude representation (distribution of activation over the number line). These different cardinal representations are directly interfaced by notation-specific input-output procedures (the verbal word frame, for example, exploits general-purpose modules for language comprehension and production), and internally, they can be transcoded by dedicated translation paths. The second premise states that different numerical procedures are tied to specific codes. Multi-digit calculations, for example, are postulated to be based on Arabic notations; verbal counting

and arithmetical fact retrieval, in contrast, rely on the verbal code. The analogue magnitude representation, finally, is required for approximate calculations and comparison tasks. Thus, the triple-code model describes numerical cognition as a layered modular architecture in which approximate magnitude codes provide the fundament for language-dependent counting abilities and symbolic manipulations processed via verbal word and Arabic digit codes.

In the following years, Dehaene and Cohen (1995, 1997) elaborated the work, introducing hypothetical anatomical substrates to some components of the triple-code model. The authors examined numerous single-case studies from patients with major hemispheric lesions in well-localized areas, hemispherectomies, or callosotomies (split-brain) who suffered from severe numerical impairments like dyslexia, alexia, or anarithmetia. Based on behavioural dissociations and deficits in relation to the specific site of a lesion they predicted the following neuronal correlates: The verbal code is processed in the left perisylvian areas, the Arabic code bilaterally activates inferior ventral occipito-temporal areas, and activity in inferior parietal areas underlies the magnitude code. In line with these neuropsychological results, early electroencephalography (EEG) and positron emission tomography (PET) imaging studies with healthy subjects reproducibly activated bilateral parietal and prefrontal regions in simple numerical comparison tasks (Dehaene, 1996; Dehaene et al., 1996).

The rise of the functional magnetic resonance imaging (fMRI) technology in the mid-90s, finally, opened completely new doors for brain research. It allowed to outline a complex neuronal *core number network* in more detail, strongly pointing (amongst others) towards the intraparietal sulcus (IPS) in posterior parietal cortex (PPC) as one of *the* integral modules for processing numerical magnitude (reviewed, for example, in Ansari, 2008; Nieder and Dehaene, 2009; Nieder, 2016a). In a quantitative meta-analysis, Arsalidou and Taylor (2011) identified and summarized brain regions concordantly activated among more than 50 fMRI studies using number tasks, confirming the inferior and superior parietal lobules (including IPS), frontal lobe areas like inferior and middle frontal gyrus of the lateral prefrontal cortex (PFC), as well as the cingulate gyrus of the medial PFC as primary number-related structures. As classic interlinked association cortices that receive highly processed input from nearly all sensory areas these regions are ideally positioned for processing numerical information.

1.2.2. Characteristics of Neurophysiological Numerical Representations

As part of an ancient and innate preverbal system (see section 1.1.2), special focus has been placed on the neurophysiological realization of the analogue magnitude code. The concept of 'numerosity' is highly abstract, generalizing across sensory modalities and spatial/temporal dimensions. Given the striking behavioural similarities when dealing with various kinds of approximate quantities (as discussed in section 1.1.1),

Dehaene (1992) assumed that different nonsymbolic numerosities converge to a common, *amodal* representation (namely the analogue magnitude) underlying one unitary neuronal basis.

Parallel to the search for neural correlates, Dehaene and Changeux (1993) developed a computational neuronal network model that accounted for elementary number processing abilities, aiming to provide both neurobiological plausibility and testable predictions. Core module of their architecture is the 'numerosity detection system' comprising several distinct layers: Visual input, in the form of 'blobs' of variable sizes at different locations on a simulated 'retina', is first normalized for size and location via an intermediate 'location map'. In the next layer, 'summation clusters' pool the activations over all positions of the location map. Much like an accumulator, discharge rates of these clusters increase or decrease monotonically and respond whenever the total activity exceeds a certain threshold. As such, they are number-sensitive, but not number-selective. Indeed, there is also neurophysiological evidence for the plausibility of such an intermediate processing stage (Romo and Salinas, 2003; Roitman et al., 2007). These summation units, in turn, are topographically linked to 'numerosity clusters' via centrally excited and laterally inhibited connections. As a result, the clusters emerging due to reinforcement-based learning are maximally activated for a preferred number and to a lesser degree also for neighbouring numbers. This non-monotonic tuning is also termed 'labelled-line coding' or 'place coding' (Nieder and Merten, 2007).

This simple multi-layer network captured the behavioural distance and size effects and, indeed, proving that also neurophysiological activity is parametrically modulated by the distance effect has nowadays become one of the most important metrics for indexing neuronal number representations.

To account for different input modalities, Dehaene and Changeux (1993) implemented an additional 'echoic auditory memory' in which auditory input was preprocessed similar to visual information and which projected also to the numerosity clusters. Probing modality-independence in living subjects, however, is much more sophisticated as numerosity is intrinsically linked with physical properties which, in turn, covary with each other. For example, a child may choose four crackers spread out on a tray rather than a pile of five crackers because the former may *appear* more numerous, given the larger area they cover. Similarly, a child's choice for four over three crackers may simply be based on the overall amount of 'edible material' rather than numerosity. Indeed, infants chose at chance when presented with one large cracker and two smaller ones of half the size (Feigenson et al., 2002). These behavioural examples already illustrate how difficult it is to disentangle numerical judgements from sensitivity to covarying sensory features. One of the key

puzzle pieces in understanding the neuronal correlates of the magnitude code was thus probing whether neurophysiological signals encode pure numerosities rather than reflecting other congruent perceptual cues from which number information is indirectly derived.

In two seminal studies, Nieder et al. (2002) and Nieder and Miller (2004) trained rhesus monkeys to perform a delayed match-to-quantity task while recording activity of single neurons in the anterior inferior temporal cortex (aITC) of the PPC and the dorsolateral prefrontal cortex (dlPFC), areas which are considered homologues of the number-responsive IPS and PFC areas in humans (Nieder, 2021). The animals had to release a lever whenever the numerosity of a dot array test stimulus matched the quantity of a previously shown sample stimulus. Abundant protocols that rigorously varied one sensory parameter after another (like covered area, density, circumference, or spatial arrangement of the dots) ensured that differential neuronal responses were attributed only to changes in quantity information of the stimuli. Indeed, a remarkably high proportion of the recorded neurons responded to the numerosity, and a substantial proportion of these cells even generalized across changes in low-level parameters of the stimuli. These *'number neurons'* responded strongest to a preferred numerosity and with progressively attenuated activity as distance from the preferred quantity increased, resulting in bell-shaped Gaussian tuning curves when considering neuronal activity as a function of sample numerosity. With increasing preferred quantity, the precision of tuning decreased (reflected in increasing bandwidth of the neural filters). Such response profiles were consistently found also for tested numerosities up to 30 (Nieder and Merten, 2007). They provided a direct neuronal correlate to the numerical distance and size effects observed in behavioural performance, and were also strikingly similar to the artificial 'number clusters' following a 'labelled-lined coding' modelled by Dehaene and Changeux (1993). The authors had thus demonstrated the existence of pure number neurons as a fundamental building block for the analogue magnitude code proposed by the triple-code model (Dehaene, 1992).

Several follow-up studies using various kinds of number stimuli during delayed match-to-quantity tasks further investigated how single neurons in PPC and PFC encode different types of quantity: Nieder et al. (2006) presented sample numerosities either as multi-dot patterns showing the number of items simultaneously, or dot by dot indicating the number of items sequentially; Nieder (2012) trained the macaques to judge the numerosities of sequentially presented auditory and visual items; and Tudusciuc and Nieder (2007, 2009) contrasted different line lengths (as continuous spatial magnitude) and multi-dot arrays (as discrete numerical quantity). All these studies revealed anatomically intermingled groups of both feature-sensitive neurons (e.g. encoding only auditory pulses, but not visual items) as well as abstract 'generalists' responsive to numerosity irrespective of modality or presentation mode. Im-

portantly, neuronal representations were surprisingly similar across areas and for different quantities, that is, bell-shaped tuning functions peaking at a preferred quantity and modulated by the numerical distance effect. Differences in response latencies and increasing insensitivity to varying sensory features (reflected in higher proportions of pure number neurons) suggested a hierarchical processing from IPS in PPC as a first cortical hub to extract numerical information to the PFC that can then process these information in an abstract, goal-directed way.

Of course, investigating number neurons in nonhuman primates is strictly limited to nonsymbolic notations. Diester and Nieder (2007), though, showed that macaques could successfully be trained to associate varying numbers with arbitrary shapes, and that these semantic associations were indeed reflected in the responses of PFC neurons, a finding that may hint at a neuronal precursor for symbolic number encoding.

Imaging studies in human subjects outline a similar picture. Piazza et al. (2004) presented their participants with short series of dot arrays and asked them to detect deviant numbers among previously shown habituation stimuli. Expectedly, behavioural responses varied systematically with distance between habituation and deviant numerosities, resulting in the characteristic bell-shaped tuning functions. In a separate adaptation fMRI experiment, subjects were instructed to simply fixate and pay attention to a passively presented sequence of dot arrays. Occasionally, deviant stimuli of different numerical value were shown amongst the habituation stimuli, again carefully controlling for other low-level visual features. The authors observed strong activations in the bilateral IPS, selectively in response to numerical changes (although numerosity was not task-relevant). Strikingly, when considering these voxel activations as a function of numerical distance, the resulting curves followed the same tuning characteristics as the behavioural ones.

This anatomical and functional homology, namely, nonlinearly compressed Gaussian tuning functions in aITC/IPS, strongly supports the notion of an innate nonverbal analogue magnitude system, inherited from our nonhuman primate ancestors.

1.2.3. The Magnitude Code: Abstract or Not?

The neuroimaging studies discussed in section 1.2.2 all used nonsymbolic stimuli to characterize the analogue magnitude code. When we talk about numbers, however, for most of us probably digits and number words come to mind first, given their omnipresence in our everyday life. The triple-code model postulated that these two forms of symbolic numbers are represented by the (visual) Arabic digit code and the (auditory) verbal word code, in separate dedicated neuronal pathways (Dehaene, 1992; Dehaene and Cohen, 1995, 1997).

There is little disagreement that notation-specific representations and processes are needed to account for the advanced spectrum of humans' numerical abilities, like arithmetic fact retrieval or sophisticated calculations (Harvey, 2016). Heated discussions (reviewed, for example, in Cohen Kadosh and Walsh, 2009), however, have been sparked on the question whether these symbolic input notations converge to the same abstract magnitude code underlying nonsymbolic number representations when we 'simply' operate with approximate quantities (as in comparison tasks).

Behaviourally, it has been argued that the numerical distance effect is highly replicable and remarkably comparable for different number notations, after 'subtracting' the time required for initial notation-dependent identification processes (Dehaene, 1996). Based on this observation, Pinel et al. (2001) introduced the term *'semantic' distance effect* to emphasize that performance depended only on numerical proximity, unrelated to the concrete appearance of the number stimuli. (That conclusions derived purely from behavioural findings can be very delusive and need to be treated with care, however, is very vividly demonstrated by the attempts to assess the subitizing phenomenon, as discussed in section 1.3.)

Also neurophysiologically, substantial evidence has been accumulated for this abstract view. Pinel et al. (2001), for example, asked subjects to decide whether a number stimulus was smaller or larger than a fixed predefined reference number. Both fMRI and high-density event-related potentials (ERPs) revealed similarly increased activations for both Arabic digits and (visually presented) number words in the 'typical' parietal number regions (amongst others along IPS and precuneus). Furthermore, in some of these areas, activity varied monotonically with numerical distance, yielding patterns directly analogous to the behavioural data.

Naccache and Dehaene (2001) used the same task design, but additionally, the numbers to be judged were preceded by subliminally presented prime stimuli. As expected, reaction times were faster when target and prime stimulus were identical (but possibly in different notations), and this well-known repetition priming effect correlated with activations of the bilateral IPS (reflected in reduced activity for identical prime-target-pairs). Analogous to Pinel et al. (2001), behavioural and neurophysiological effects were independent of both target and prime notations. Thus, their study provided evidence that numbers were also genuinely processed when perceived unconsciously.

Eger et al. (2003), finally, showed their subjects interleaved sequences of numbers, letters, and colours, presented visually and auditorily, and asked their subjects to report the occurrence of target items that were identical across modalities within each category. This task design was very conclusive in two aspects: First, presenting stimuli in different sensory modalities allowed the search for supramodal number representations, beyond the visual range. Second, this paradigm did not only prevent being

biased towards any category or modality, but more intriguingly, number information was essentially irrelevant, thereby circumventing potentially confounding task-driven effects that might explain the activation of magnitude representations. Indeed, only the IPS showed activations common to both modalities and specific to numbers. Further control tasks confirmed that these implicit supramodal number representations were colocalized with those activated during explicit numerical processing.

Taken together, these studies consistently demonstrated that the IPS is activated similarly by symbolic number stimuli of different notations and/or sensory modalities, with activations modulated by the numerical distance effect and irrespective of whether number information is processed explicitly, implicitly or even perceived unconsciously.

Interestingly, these findings led to the conclusion that numbers had been converted into a common semantic quantity code. They were all based on the exploration of symbolic stimuli, though, and neither of these studies had actually contrasted symbolic and nonsymbolic number formats directly, thus leaving open the possibility of anatomically overlapping, but functionally segregated systems.

To clarify this issue, Fias et al. (2003), for example, presented their subjects with pairs of angles and lines of different magnitudes as well as Arabic digits, and asked them to indicate the largest quantity. PET recordings revealed a region in left IPS that was specifically activated by the number, irrespective of stimulus format, thus providing direct evidence for common activations. Recording ERPs during a numerical comparison task, Libertus et al. (2007) showed that the electrophysiological correlates of the numerical distance effect during specific ERP components were identical for symbolic Arabic digits and nonsymbolic random dot arrays (at least, when choosing numbers well beyond the subitizing range).

Finally, using an fMRI adaptation paradigm, Piazza et al. (2007) asked their subjects to passively fixate on a screen and pay close attention to the quantities presented as dot arrays or Arabic digits (thereby avoiding any decision and response confounds). During an initial adaptation period, subjects were shown a series of approximate quantities from within a very limited range (e.g. 17, 18, 19) of the same format. Afterwards, deviant values of close (e.g. 20) or far (e.g. 50) numerical distance were suddenly interspersed, with or without a concomitant change in format. During adaptation, IPS activity decreased continuously, but suddenly recovered once the deviant stimulus was presented, indicating that the neuronal populations were sensitive to the altered numerosity. These rebound effects varied in a distance-dependent fashion, but, critically, were independent of format and *changes* in format.

Findings like these further corroborated the idea of an abstract magnitude representation. Unfortunately, due to their large-scale spatial resolution, recording methods like fMRI or EEG cannot disentangle whether the observed common effects

for nonsymbolic and symbolic stimuli result from truly abstract number cells or arise due to the combined recordings of anatomically intermingled subpopulations of format-specific number neurons (as also found in monkeys; see section 1.2.2).

Indeed, there is also convincing evidence that quantities of different formats or notations are encoded differently. Already Buckley and Gillman (1974) reported that the distance effect was less pronounced for the Arabic digits than for the nonsymbolic stimuli. Similarly, behavioural studies with both children and adults (Holloway and Ansari, 2009; Maloney et al., 2010) provided evidence that the numerical distance effects elicited by nonsymbolic and symbolic stimuli, respectively, are scarcely correlated, thus challenging the idea of a universal 'semantic distance effect' put forward by Pinel et al. (2001). The authors argued that this lack of correlation may instead imply different underlying neurophysiological processes.

Indeed, Cohen Kadosh et al. (2007) observed strong hemispheric differences in parietal activation during cross-notational adaptation to numerals and number words (notably, again two symbolic notations). Specifically, they reported adaptation in the left IPS regardless of the stimulus notation, arguing for notation-independent representations; in the right IPS, in contrast, notation and quantity interacted significantly, yielding adaptation effects only for Arabic numerals.

Interestingly, also Piazza et al. (2007) (who had used dot arrays and Arabic numerals) reported and discussed hemispheric differences, specifically, in the right IPS distance effects were independent of changes in format from target to deviant numbers, in contrast to format-sensitive distance effects in the left IPS suggesting higher precision of magnitude coding. Conciliating these seemingly contradictory findings, Ansari (2007) argued that both studies hint at a specialized role of the left IPS in the representation of encultured number symbols (numerals and number words, as shown by Cohen Kadosh et al., 2007), also in terms of coding precision (as shown by Piazza et al., 2007) – without ruling out the possibility of abstract number processing –, thus shifting the focus away from the dichotomous question of *whether* to *where* and *how* abstract and nonabstract number representations (co)exist.

All the studies discussed here addressed the question of representational overlap between symbolic and nonsymbolic quantities by searching for commonly activated regions whose activity was comparably (or differentially) modulated by different notations. Similar responses as evidence for abstractness, however, need to be considered with caution as they do not necessarily imply shared neuronal representations. To directly probe the numerical coding by distributed activity patterns in IPS, Eger et al. (2009) recorded high-resolution fMRI data while subjects performed a delayed match-to-numerosity task, presenting dot arrays of different dot size and luminance, and Arabic digits as stimuli. Via multivariate pattern analysis (MVPA), the authors

identified voxels commonly activated by all stimuli and used these data to train a classifier to discriminate between different stimulus conditions. Within each format, number information was significantly decoded, that is, a classifier trained on dot arrays of different numerosities could reliably predict the numerosity of novel dot arrays, and analogously for Arabic digits, yielding, however, more robust effects for the nonsymbolic stimuli. Surprisingly, generalization performance across formats was strongly asymmetric, that is, a classifier trained only on Arabic digits succeeded in decoding number information of nonsymbolic dot arrays, but accuracy for the reversed direction (i.e. trained on dot arrays and tested on Arabic digits) was at chance level. Interestingly, the classifiers successfully generalized across low-level features (i.e. a classifier trained only on data from numbers of constant dot size could accurately discriminate data from numbers of constant luminance, and vice versa), and failed to discriminate differences in luminance or dot size when trained on a specific number. Discrimination of stimulus format (dot array versus Arabic digit), however, was highly accurate.

These findings confirmed once more, that the neuronal number codes in IPS reflect primarily quantity rather than secondary low-level features, but revealed again a striking sensitivity to the specific format of number. Specifically, the asymmetry in generalization fits nicely with the idea of more narrowly tuned symbolic representations in addition to the evolutionary older broadly tuned nonsymbolic representations (as observed already by Piazza et al., 2007).

Such differences in coding precision are indeed in line with predictions of the influential neuronal network model by Verguts and Fias (2004) that explicitly simulated the linkage between nonsymbolic and symbolic quantities into a common representation. Their model starts from the presumed results of any sensory-specific preprocessing stages, thus boiling down input to the key properties of numbers of different formats. Centerpiece of the model is the 'number field' that receives nonsymbolic input via a 'summation field' with similar properties as the analogue proposed by Dehaene and Changeux (1993) (see section 1.2.2), i.e. the amount of activated nodes corresponds to a particular quantity. In parallel, symbolic input is provided by the 'symbol field' in form of an arbitrarily chosen single unit representing that very quantity, thereby accounting for the properties that symbols are related arbitrarily to each other and their physical appearance bears no numerical information (see section 1.1.3). During unsupervised learning, nonsymbolic and symbolic stimuli were presented together. Afterwards, the model was given only nonsymbolic input. Indeed, number-selective output units had emerged that were characterized by the same critical tuning properties observed in the number neurons of macaques Nieder et al. (2002) and predicted also by the model of Dehaene and Changeux (1993), i.e. nonmonotonic filter functions indicative for labelled-line coding

with increasing bandwidths. Critically, the activations of the very same units showed remarkable differences when stimulated with symbolic input only. They maintained their number preference, i.e. units preferring a particular numerosity were also maximally activated by the corresponding symbol. Their bandwidths, however, were notably smaller resulting in more 'peaked' curves that were hardly skewed. In other words, symbolic representations were more precisely, but still partially inherited some properties of the primordial nonsymbolic ones.

Altogether, both neurophysiological recordings and simulations with computational models suggest that format dependency does not pose a conceptual problem to number coding, and that a neuronal system that is originally devoted to process the evolutionary older nonsymbolic quantities can learn to represent also arbitrary number symbols. The properties of the involved neuronal populations, though, may differ depending on the input format.

1.2.4. A More General Sense of Number

Over the last few decades, the search for neuronal correlates of numerical representations has spawned numerous many-faceted and most diverse findings that further extended the idea of an innate number sense.

Based on the observation that numerosity was strongly susceptible to adaptation (which is assumed to be a clear signature for perceptual mechanisms; Wark et al., 2007), Burr and Ross (2008) suggested that numerosity estimation might already take place in the visual system, putting forward the notion of a 'visual sense of number'. Indeed, there is evidence that both humans (Cicchini et al., 2016; Park et al., 2016) and monkeys (Viswanathan and Nieder, 2013) can perceive and encode nonsymbolic quantities spontaneously, even in the absence of explicit number-related task demands, and also modelling studies (Stoianov and Zorzi, 2012; Nasr et al., 2019) show that tuned number neurons emerge spontaneously in artificial neuronal networks merely trained on visual object recognition.

Furthermore, Harvey et al. (2013) showed that neuronal populations in the parietal cortex were topographically organized according to numerosity. Such forms of representations are commonly found throughout the sensory and motor cortices (Kaas, 1997), but apply also to higher-order cognitive functions (Thivierge and Marcus, 2007; Anderson et al., 2010) (probably due to their computational efficiency; Chen et al., 2006). Later, Harvey and Dumoulin (2017) identified an entire network of such topographic numerosity maps throughout the human association cortices, overlapping with other brain areas involved, for example, in visual motion processing, object recognition, attentional control, or decision making. On a single-cell level, studies with macaques revealed causal links between the activity of cells in sensorimotor ar-

eas and numerical information about self-generated motor actions (Sawamura et al., 2002, 2010).

Without questioning the abstracted representations of magnitude in the higher-order association areas, all these findings led to the idea that the 'number sense' comprises a complex, extensive 'sensorimotor numerosity system' that links perception, higher cognitive functions and action (Anobile et al., 2021).

1.3. Subitizing in Enumeration Processes

Over millennia, different civilizations developed different number symbols and enumeration systems. A closer look at ancient numerical notations, however, reveals a striking parallel across many (if not all) societies: The first three (or sometimes four) numbers were almost always denoted by sets of identical tokens, e.g. Roman I, II, III or Mayan •, ••, •••, ••••; followed by often arbitrary symbols for larger numbers, e.g. Roman IV, V or Mayan ¯ (Ifrah, 1981). This is, however, by far no coincidence but can be explained by the peculiarities of our cognitive system: While large nonsymbolic quantities need to be counted slowly and effortfully, humans can identify up to three or four objects rapidly and accurately 'at a single glance'.

On a scientific level, this observation has occupied the minds of philosophers and psychologists for over a hundred years. Already in 1871, Jevons (1871) reported that he could judge the number of beans casually thrown into a box accurately for only up to 3 or 4 pieces, and that he made increasingly more errors for larger amounts of beans. Kaufman et al. (1949) observed the same behavioural dichotomy between small and large quantities when he asked subjects to count the number of briefly presented, randomly arranged fields of dots. He termed this phenomenon *'subitizing'*, derived from the Latin adjective 'subitus' (= sudden) and the medieval Latin verb 'subitare' (= 'to arrive suddenly'), to describe the 'rapid, confident, and accurate report of the numerosity of arrays of elements presented for short durations' (Mandler and Shebo, 1982, p. 1).

Countless subsequent studies consistently confirmed this dichotomous behavioural response pattern, interestingly, not only for visually presented quantities, but also for stimuli from the auditory (ten Hoopen and Vos, 1979; Repp, 2007; Camos and Tillmann, 2008; Anobile et al., 2019) and tactile (Riggs et al., 2006; Plaisier et al., 2009) domain. The debate on what type of process subitizing actually is, however, is still not resolved.

1.3.1. Subitizing versus Estimation: One or Two Representational Systems?

Strongly influenced by a developmental psychological perspective, the theory of *'core knowledge'* (Spelke and Kinzler, 2007) proposes that humans are by birth endowed with a few cognitive systems upon which all knowledge acquired in life is built.

One of these systems is the ANS that serves to represent sets and their numerical relationships of ordering, addition and subtraction (see section 1.1.1). A second one is the *'object tracking system'* (OTS) for object representations, following the low-level spatio-temporal principles of cohesion (objects are connected as wholes), continuity (they move on connected, unobstructed paths) and contact (they influence each others' motion only when they touch). This system can be easily influenced by object-directed attention and is as such assumed to be strictly limited in capacity (van Marle and Scholl, 2003; Marino and Scholl, 2005).

Developmental studies revealed that infants' performance in various numerical tasks diverged drastically from the ratio-dependence characteristic for the ANS (see also section 1.1.2), but rather depended strongly on the absolute number of objects presented (Feigenson et al., 2002; Feigenson and Carey, 2003). As the observed upper-bound of 3 nicely coincided with the limits of the subitizing range in adults (Kaufman et al., 1949; Mandler and Shebo, 1982), the authors suggested that subitizing reflected not the ANS but rather the OTS which would serve as a different *enumeration* system, complementing the ANS and explicitly dedicated to the precise representation of small numbers of distinct individuals (see Feigenson et al., 2004, for a review). In a similar vein, Carey (2002) argued that the ANS might not be suited to represent natural numbers due to its imprecision, thus, the first meaning of numerals would be provided by the OTS.

No doubt, children start with acquiring the semantic meaning of 'one', then 'two' and 'three', and only once this is mastered, they move on to all other numbers in their count list (Wynn, 1990). However, while there is substantial evidence for a foundational role of the ANS in the acquisition of symbolic numbers and arithmetic (see also section 1.1.2), the link between OTS/subitizing and maths seems to be somewhat variable (Piazza, 2010; Anobile et al., 2019).

In fact, there is no reasonable explanation for why humans should have evolved two qualitatively different systems *specialized for numbers*. More recently, researchers convincingly argued that the observed behavioural discontinuities between subitizing and estimation are simply explainable as a consequence of the nonlinear compression of the ANS' mental number line on which small and large numbers are continuously represented with increasingly broader, less overlapping tuning functions (see also section 1.1.1).

According to signal detection theory (Green and Swets, 1966; Stanislaw and Todorov, 1999), discriminability between two adjacent numbers (as quantified by the index d') depends on the amount of overlap between the tuning curves. A modelling study by Tsouli et al. (2022) showed that this relationship does not only explain the characteristic size effect. Moreover, with d' being very high for up to four items, discriminability performance in the subitizing range would simply be at ceiling and

therefore error-free.

Similarly, connecting number psychophysics and an information-theoretic modelling approach, Cheyette and Piantadosi (2020) argued that different response distributions for small and large numbers would naturally emerge in a single system that efficiently exploits the nonuniformity of the 'need probability' of number (Anderson and Schooler, 1991) (i.e. the fact that small numbers are more often encountered and used than larger ones; Dehaene and Mehler, 1992) to account for its limited informational capacities, thereby meeting the constraint that internal representations should not be more accurate than the perceptually provided information (Gallistel, 2017) and optimizing the trade-off between processing costs and the benefits of accurate perception.

1.3.2. The Role of Nonnumerical Mechanisms in Subitizing

Not denying the obvious differences between subitizing and estimation, recent research seems to consider subitizing less a separate number system but rather a phenomenon manifesting within a distributed, quite general enumeration system (see section 1.2.4) as a result of the complex interplay of many nonnumerical processes such as attention, working memory and object tracking.

There is consistent evidence that – unlike estimation (see, however, Pomè et al., 2020, for relativizing evidence) – subitizing strongly depends on attentional resources, as typically manipulated in dual-task experiments (Railo et al., 2008; Burr et al., 2010; Piazza et al., 2011). Burr et al. (2010), for example, reported small Weber fractions (indicative for highly precise, nearly error-free responses) for small numbers in single-task conditions. Under conditions of high attentional demands (that did not allow the subjects to pay full attention to the numerosity stimuli), in contrast, this characteristic signature vanished and seemed to make room for the ANS, as reflected in Weber fractions approaching those observed also in the estimation range.

Similarly, Piazza et al. (2011) showed that individual differences in the subitizing capacity varied systematically with differences in visual working memory capacity (which is also limited to up to four items; Luck and Vogel, 1997), but not with differences in large number estimation precision (Revkin et al., 2008). They argued that these findings hint at a domain-general mechanism for parallel object individuation (like the OTS).

Indeed, the strong impact of such visuo-spatial aspects is also reflected in the observation that subitizing effects are very pronounced when the individual items are presented simultaneously, in different spatial locations, and pop-out from the background, and are even more enhanced for canonical configurations, (e.g. dice patterns) (Mandler and Shebo, 1982; Krajcsi et al., 2013). They vanish, however, for complex, less well-specified spatial arrangements, for example, when objects are tangled or shown as transparent, but overlapping shapes, or when pop-out is prevented (e.g. de-

tecting 'O's among 'Q's), thus involving probably also pre-attentive processes (Trick and Pylyshyn, 1994).

Finally, subitizing limits seem to vary between different input modalities (visual versus auditory), and within the visual domain also between simultaneously versus sequentially presented stimuli, indicating that subitizing in itself is not a unitary thing but may rather be subserved by several separate processes (Anobile et al., 2019).

All these examples show that subitizing taps many cognitive mechanisms to which estimation seems to be less susceptible – disentangling the individual contributions of the many highly interacting processes, however, turns out to be very tricky.

1.3.3. Neuronal Correlates for Subitizing

What is already highly complicated at a behavioural and conceptual level proves to be even more intricate at a neurophysiological level.

In recent years, evidence seems to have converged to the view that the fronto-parietal 'core' number network hosts representations of both small and large numerosities alike (Piazza et al., 2002; Cai et al., 2021; Fornaciai and Park, 2021; Tsouli et al., 2022) (see, however, Hyde and Spelke, 2009, for an opposing position). Based on the work of Harvey et al. (2013), who had shown that neuronal populations in the human parietal cortex were topographically organized according to numerosity, Cai et al. (2021) provided evidence that small and large numerosities were continuously encoded within the same numerosity map that was characterized by cortical magnification (i.e. more cortical surface area was devoted to smaller numerosities) and a tuning precision that decreased systematically with preferred numerosities. Tsouli et al. (2022) argued that these differences in tuning properties and distinct proportions of neurons preferring different numbers suffice to explain the dichotomous behavioural phenomena of subitizing and estimation.

In a similar vein, and in the light of the undeniable link to attentional and memory processes, modern computational architectures, that combine important features of existing number- and attention-models, show that behaviourally distinct ranges of numbers can emerge also within the same neuronal network by flexibly adapting its internal parameters (Stoianov and Zorzi, 2012; Sengupta et al., 2014; Knops et al., 2014).

The neuronal network architecture developed by Dehaene and Changeux (1993) is one of the earliest computational models of numerosity. Though not addressing the subitizing/estimation dispute, it satisfactorily explained many characteristic behavioural findings (following Weber's law), and many of its predictions concurred with subsequently revealed neurophysiological observations – making it unquestionably also one of the most influential models. Their work, however, rested on some critical *a priori* assumptions that were hard-coded in the architecture (see section 1.2.2).

Most importantly, object normalization for size and position was accomplished by a simple inhibitory surround mechanism (i.e. recurrent self-excitation in the 'on-centre' and lateral inhibition by all other nodes in the 'off-surround') within the intermediate 'location map', thereby effectively creating spatial segregation. Almost twenty years later, Stoianov and Zorzi (2012) observed that within a simple 'deep' neural network, that was merely trained on sensory data but not on numerosity discrimination, numerosity-selective nodes emerged spontaneously, and that these nodes were indeed characterized by surround inhibition profiles.

Interestingly, detached from numerical concepts and exclusively focusing on spatial attention, Roggeman et al. (2010) developed a computational model for spatial 'saliency maps' (i.e. topographic maps in which neuronal activations represent salient objects that stand out the most from the environment; Koch and Ullman, 1985) – which are thought to be the basis of attention and working memory –, that rested also on object segregation via surround inhibition. In their attempt to investigate the influence of task demands on the activations of the saliency map, they showed that the level of the lateral inhibition between the network nodes determined the capacity limit of the saliency map.

Building on the findings of Roggeman et al. (2010) and Stoianov and Zorzi (2012), finally, Sengupta et al. (2014) explored a simple on-centre off-surround neural network with varying degrees of surround inhibition. Directly addressing the subitizing/estimation debate, they showed that the same network could account for both small and large number processings depending on the inhibition strength; high levels of inhibition could reliably encode small quantities, whereas networks with moderate levels of inhibition accounted better for larger numerosities within the estimation range. What may regulate the inhibition strength – pre-attentive mechanisms or rather task-dependent top-down modulation – remains an open question, though.

Of course, such artificial neuronal network architectures are always highly simplistic abstractions of the biological brain, and are as such inherently limited in their explanatory power. All these models, however, hint at surround inhibition, which is indeed a basic neuronal circuit operation known to be involved in shaping the tuning of selective neurons (Hartline et al., 1956; Isaacson and Scanziani, 2011), as a critical neuronal building block leading to the differences between subitizing and estimation.

1.4. Mathematics in the Brain

Unquestionably, even the best mathematicians started their career with learning how to count and how to perform simple operations like addition, subtraction, multiplication and division. To investigate the origins and fundamental building blocks of mathematical concepts, most studies therefore focus on tasks requiring only basic arithmetic as taught already at primary school. Calculation requires the identification of two or more numbers, online maintenance of the numerical values in working

memory, and finally the subsequent modification based on the operational function – and disentangling the underlying cognitive processes, memory representations, and mental components has proven to be a rather intricate challenge.

1.4.1. Core Brain Regions for Mental Arithmetic

Abundant behavioural research over the last decades provided multiple evidence for basically two opposing key strategies to solve simple arithmetic problems, that is, direct retrieval of rote numerical facts from declarative memory (i.e. knowing by heart that 9+4 = 13) (Ashcraft, 1992, 1995), contrasting alternative calculation procedures, for example, counting (e.g. 9... 10..11..12..13), transformation/decomposition (e.g. 9+4 = 9+1+3 = 10+3 = 13), or inversion (i.e. exploiting that subtraction is the opposite of addition and, thus, 13–4 = 9 corresponds to 9+4 = 13) (LeFevre et al., 1996; Hecht, 1999; Campbell and Timm, 2000; Campbell, 2008). While addition and multiplication seem to be primarily solved via retrieval, subtraction and division seem to rely more on alternative strategies (Campbell and Xue, 2001).

In the triple-code model (see also section 1.2.1), the distinction between different arithmetic operations manifested in two different proposed pathways (Dehaene and Cohen, 1997): Addition and multiplication tables, stored as verbally encoded rote numerical facts, were assumed to be accessed via a 'direct asemantic route', recruiting the left cortico-subcortical loop through basal ganglia and thalamus. Subtractions, in contrast, for which rote memory is usually unavailable would be processed via an 'indirect semantic route' that transcodes operands into quantities for semantically meaningful manipulations, eliciting activity in the bilateral inferior parietal cortex and left perisylvian language network. Finally, the prefrontal cortex was proposed to be involved in strategy choice and planning.

Since the initial conception of the triple-code model back in the 1990s, whose suggested neurophysiological correlates stemmed primarily from single-case lesion studies, substantial research has further deepened our understanding of the neurophysiological processes underlying mental arithmetic. Countless fMRI studies with healthy subjects (see Arsalidou and Taylor, 2011; Menon, 2015, for a review) as well as studies using intracranial electrocorticography (ECoG) in epileptic patients (Allison et al., 1999; Shum et al., 2013; Daitch et al., 2016; Hermes et al., 2017; Pinheiro-Chagas et al., 2018) confirmed that arithmetic operations concordantly activated regions similar to those in pure number tasks, most importantly, ventral occipito-temporal areas (associated primarily with the visual recognition of the stimuli) as well as the well-known dorsal aspects of the posterior parietal cortex (that is, horizontal IPS, angular gyrus and supramarginal gyrus). In line with such correlational evidence, direct cortical electrostimulation (DCE) in patients undergoing awake surgery for removal of pathological tissues revealed localized small cortical areas in the 'classical' core number network in which DCE stimulation provoked a temporary disruption

of calculation performance (Roux et al., 2009; Della Puppa et al., 2013; Semenza et al., 2017), and many of the patients in which these calculation-related areas could not be spared for oncological reasons developed significant acalculia symptoms postoperatively (Roux et al., 2009), confirming direct causal involvement of these regions in mental arithmetic.

Many of these studies reported that calculation-responsive sites were differentially activated (like the IPS) or deactivated (especially the angular gyrus) for addition, multiplication, subtraction and division (e.g. Rosenberg-Lee et al., 2011, comparing all four operations). This functional heterogeneity, however, seems to indicate *operation*-specificity only insofar as it reflected the variable and different reliance on general processing *strategies* typically associated with specific arithmetic operations (Grabner et al., 2009; Rosenberg-Lee et al., 2011; Tschentscher and Hauk, 2014; Menon, 2015). And indeed, strategy use differs widely across individuals and also within operations (Campbell and Xue, 2001, for example, observed retrieval use only for 88 % of their simple addition problems according to subjects' verbal reports) – averaging data over strategies, however, was proven to result in misleading conclusions (Siegler, 1987). Thus, rather than contrasting specific operations, more recent studies focussed on the different strategies used for mathematical problem solving (see also section 1.4.2).

Unlike number tasks that comprise counting, ranking, or comparing quantities, calculation tasks require the identification and memorization of (at least) two numbers and their subsequent modification based on the operational function. Pronounced activations of more prefrontal areas (extending those activated also during number tasks) suggest that solving arithmetic problems requires more cognitive resources like working memory and attention (e.g. Gruber et al., 2001; Arsalidou and Taylor, 2011; Menon, 2015). In line with this idea, prefrontal contributions were noticeably affected by task difficulty (Menon et al., 2000; Fehr et al., 2007), for example, number of steps required, or single- vs. two-digit calculations, an important factor that modulates also behavioural performance (e.g. Ashcraft, 1992; Agostino et al., 2010). Furthermore, and consistent with previous studies focussing on reasoning and working memory tasks (Christoff and Gabrieli, 2000; Owen et al., 2005), evidence suggests that the prefrontal cortex is involved whenever the brain has to deal with more than one item and/or when processing involves manipulation of information according to *rules*, i.e. conditional 'if-then' statements, as befitting its general role in executive control (Miller and Cohen, 2001; Fuster, 2015).

1.4.2. The Medial Temporal Lobe (MTL) in Mathematical Cognition

For quite a long time, research on mental arithmetic has focused on the interplay of the 'domain-specific' intuitive number sense providing semantic representations of

quantities derived from all kinds of sensory inputs, and executive control processes guiding planning and goal-directed decisions based on the extracted numerical information. These processes residing in the core network of posterior parietal, occipito-temporal and prefrontal areas (as outlined in the previous sections), however, are but two building blocks necessary for arithmetic. Considering the importance of different strategies used for mathematical problem solving, over the last years, the focus of attention has shifted also to 'domain-general' processes relevant for various academic skills and learning in general (Menon, 2015, 2016; Peters and De Smedt, 2018). Episodic and semantic long-term memory (LTM) systems, for example, are required for the formation and retrieval of rote arithmetic facts; alternative strategies, in contrast, require the recruitment of procedural and working memory (WM) systems for memorization, sequencing and manipulation of operands.

The medial temporal lobe (MTL) – comprising parahippocampal cortex (PHC), entorhinal cortex (EC), hippocampus (HIPP), and amygdala (AMY) – is well-known for its role in declarative memory (Squire and Zola-Morgan, 1991; Tulving and Markowitsch, 1998), and is characterized by unique brain responses and neuroanatomical connections (Squire et al., 2004). Recordings in human epileptic patients implanted with depth electrodes for intracranial EEG have revealed that single units of the MTL respond selectively to images from various categories, like faces, animals, objects or scenes (Fried et al., 1997; Kreiman et al., 2000; Mormann et al., 2011), often showing a high degree of visual, e.g. size or viewing angle (Quian Quiroga et al., 2005), but also multimodal invariance, e.g. image, spoken and written name of the stimuli (Quian Quiroga et al., 2009). The sparse, explicit and abstract representations of these *'concept cells'* are nowadays considered crucial building blocks of declarative memory functions (Quian Quiroga, 2012). Importantly, the highly associative brain areas are directly and reciprocally connected with the frontal number network (Goldman-Rakic et al., 1984). These hippocampal-neocortical pathways are involved in memory storage by delivering highly processed information from dlPFC to hippocampus; return projections, in turn, provide access to the highly associative memories stored in the intrinsic circuits of the hippocampus.

Memory systems play the most pivotal role during ontogenetic development and maturation in childhood, when problem solving skills are still unfolding. Over the last years, it was shown that children indeed rely also on areas not previously recognized in the mental 'maths network' (see Arsalidou et al., 2018, for a review). The MTL, in particular, appears to be critically involved during the development of arithmetic fact knowledge (Menon, 2016).

The acquisition of arithmetic competences is (amongst other things) characterized by developmental shifts in the strategies used (Siegler, 1996; Jordan et al., 2003; Geary, 2011): Already before the start of formal education, children use counting to solve

simple addition problems, often executed with the help of, for example, the fingers. Yet progressively, the efficiency of this strategy is increased by moving to verbal counting (without external manipulatives), and is further refined by starting to count from the cardinal value of the first (e.g. 2+3 = 2... 3..4..5) or larger (e.g. 3... 4..5) number instead of counting all elements of both numbers (e.g. 1..2... 3..4..5) (Geary et al., 1992). In the course of maturation of general problem solving approaches, finally, children move on to more advanced and efficient memory-based strategies (Peters and De Smedt, 2018).

Strategy use and its gradual transition in behaviour is also reflected in developmental changes of the 'maths network', both in terms of function, connectivity and structure (Peters and De Smedt, 2018). De Smedt et al. (2011), for example, reported greater hippocampal activations indicative of fact retrieval only when children solved simple addition problems; difficult calculations and subtractions that are less well rehearsed and more difficult to memorize, in contrast, recruited the fronto-parietal network suggesting a stronger influence of procedural strategies. In line with this findings, Qin et al. (2014), who accompanied 7–9-year old children in a longitudinal fMRI study over a one-year period, observed that the shift from counting to arithmetic fact retrieval was characterized by an increased engagement of hippocampal areas and decreased activations of prefrontal-parietal regions. Strongest predictor for individual improvements and gains in fact-retrieval fluency, however, was the increased connectivity in hippocampal-neocortical circuits (Cho et al., 2012; Qin et al., 2014).

Similarly, Supekar et al. (2013) showed that the success of an intensive maths tutoring program with 3rd-graders was strongly correlated with individual differences in grey matter volume of HIPP and its connectivity with PFC and basal ganglia measured before tutoring. In contrast, comparisons between typically developing children and ones with developmental dyscalculia, who typically have difficulties retrieving arithmetic facts from memory (Geary et al., 2004), indicated decreased grey matter volume in several brain regions, including the PHC (Kucian et al., 2006; Rykhlevskaia et al., 2009).

Taken together, although studies in adults scarcely report hippocampal engagement during arithmetic tasks (which may be explained by a stronger reliance on neocortical memory systems in the course of memory consolidation, as proposed by Eichenbaum et al., 2007), behavioural and neuroimaging data of children provide evidence for the causal role that the MTL plays in mathematical cognition.

1.4.3. Monkey Single Neurons Encoding Abstract Rules

Mental arithmetic is undoubtedly a hallmark of our scientifically advanced human culture. Operating on most abstract categories and principles, exact calculations are closely linked to the mastery of symbolic number representations and the understanding of rules associated with mathematical signs.

Interestingly, however, most studies using arithmetic tasks do not address the 'rule-component' of calculations explicitly. Of course, contrasting addition and multiplication asks for the application of two different rules. However, in most paradigms all elements of the calculation were presented simultaneously (e.g. 3+5 with or without a proposed result requesting verification or computation), which prevents disentangling rule-specific activity from other processes relevant for calculation. Daitch et al. (2016) were one of the few who used also element-wise presented calculations while recording ECoG in lateral parietal and ventral temporal cortex of human subjects. And indeed, they reported differential involvement of some regions in different stages of numerical processing; rule-specific activations, however, were not explicitly analysed.

Mathematical and numerical capabilities of nonhuman primates, in contrast, are strictly confined to the nonsymbolic domain of the ANS. And although its imprecise numerical representations may suffice for approximate calculations (see section 1.1.2), so far, the neuronal correlates of addition or subtraction have not been studied in monkeys. However, also the standard numerical comparison task commonly used in studies with human subjects calls for the application of a mathematical rule: In order to decide, for example, 'is the sample number smaller or larger than 65?' (the task instruction given by Pinel et al., 2001), the presented quantity X (whether exact or approximate) first needs to be sorted according to its ordinal relationship relative to the reference number: X<65 or X>65. In a second step, conditional 'if-then' statements then determine the requested goal-directed behaviour: '*if* X<65, *then* press the right key', and '*if* X>65, *then* press the left key'. As already demonstrated by Brannon and Terrace (1998) (see also section 1.1.2), these simple ordinal *'greater/less than' rules* underlying number comparison tasks can also be mastered by nonhuman primates.

Early studies in rhesus monkeys (White and Wise, 1999; Asaad et al., 2000) have shown that single neurons in the PFC are not only sensitive to sensory information like stimulus appearance and location, but that their activations can also reflect information about the task being performed. In their experiments, stimuli and trial events were identical across tasks, but each task asked for the application of a different rule. Which rule, however, was not (in Asaad et al., 2000) or only indirectly (in White and Wise, 1999) conveyed, and the monkeys had to deduce this information from the fact that within a block of trials, the rule was always the same.

To address rule-selectivity more explicitly, Wallis et al. (2001) thus trained macaques on a delayed match/nonmatch-to-sample task that asked for different behavioural responses depending on two abstract rules; the 'match' rule required the animals to respond when sample and test images were identical, the 'nonmatch' rule required a response when the stimuli were different. In their paradigm, the current rule varied randomly from trial to trial, and was indicated explicitly by a cue presented

simultaneously with the sample. Two distinct cues for each rule, respectively, allowed disentangling rule- and sensory-driven activity. Indeed, the authors recorded a substantial proportion of PFC neurons that encoded these abstract 'same/different' rules, irrespective of the concrete rule cues applied.

Bongard and Nieder (2010), finally, bridged the gap between rule-selectivity and mathematical cognition, training rhesus monkeys on a rule-based numerical discrimination task (similar to Wallis et al., 2001) in which the animals had to compare numerosities presented as random dot patterns, and to switch flexibly between two mathematical rules; the 'greater than' rule required the monkeys to respond when the test display showed more dots than the sample stimulus, vice versa, the 'less than' rule required a response when the test stimulus contained less dots than the sample. Again, two distinct cues from different sensory modalities were used for each rule, respectively. Short delay periods between the presentation of sample, rule cue and test stimulus, respectively, enabled the authors to analyse also activity related to working memory processes. Behavioural performance for the task was not only well above chance level, the animals were also able to immediately generalize the two rules to previously unseen quantities, confirming that they had learned the abstract mathematical principles. These were also reflected in rule-selective activations for a substantial proportion of single neurons in the dlPFC that responded preferably to either of the two rules, insensitive to sensory features of the rule cues. These activations were directly correlated with behavioural responses of the monkeys: If the animals made errors, discharge rates for the preferred rule were also markedly decreased.

As discussed in section 1.2.2, quantity information is encoded by intermingled neuronal subpopulations of both 'specialists' that encode only one type of magnitude, and 'generalists' encoding quantity as an abstract concept. Similarly, Eiselt and Nieder (2013) showed that also the majority of rule-selective cells distinguished between different magnitude types (i.e. they encoded 'greater/less than' rules either for continuous line lengths or discrete dot numerosities), in addition to a significant proportion of 'rule generalists' that encoded the overarching abstract concept 'magnitude rules'.

Finally, rule-selective neurons were not confined to prefrontal areas but were also observed in monkeys' premotor and posterior parietal cortices (Wallis and Miller, 2003; Stoet and Snyder, 2004; Vallentin et al., 2012). Although differences in cell frequencies and coding properties provide evidence for a hierarchical and specialized involvement of different areas in task execution, this finding indicates that all task-relevant features (like stimulus identity, current rule, and behavioural response) may be represented to a greater or lesser degree by all brain regions along the processing pipeline from perception to goal-directed action.

1.5. Working Memory in the Medial Temporal Lobe

Shifting to a multisystem approach for arithmetic problem solving, recent research also emphasized the role of working memory systems in mathematical cognition. Traditionally, WM is defined by its functional properties as the ability to keep a restricted number of events or stimuli 'in mind' for short periods of time (usually on the scale of seconds) until a response is required, even when the stimuli are no longer physically present (Baddeley and Hitch, 1974; Baddeley, 2003). As such, it is typically considered a complement to long-term memory which can durably store vast amounts of information (Atkinson and Shiffrin, 1968; Cowan, 1988).

Multiple strands of evidence, however, are currently revolutionizing the conventional conceptualization of working memory and its underlying neuronal correlates (e.g. Kamiński and Rutishauser, 2020; Rose, 2020; Beukers et al., 2021; Foster et al., in press).

1.5.1. Different Neuronal Codes for Memories

Neurophysiologically, it was assumed for a long time that the WM system was anchored primarily in parietal-frontal circuits (Curtis and D'Esposito, 2004; Funahashi, 2017), and that *sustained* representations in WM were indexed by single neurons with persistently increased firing patterns throughout delay periods between stimulus presentation and response (Fuster, 1971; Golman-Rakic, 1995). Such task-specific delay activations in WM have been reported for many kinds of memory items, including the rule-selective single neurons found by Wallis et al. (2001) and Bongard and Nieder (2010) (see section 1.4.3).

A growing body of evidence shows that the WM system is not restricted to frontal areas but operates across a widely distributed cortical network. The fact that hippocampus and surrounding MTL areas are critical for LTM is firmly established (Squire and Zola-Morgan, 1991; Tulving and Markowitsch, 1998; Squire et al., 2004). More recently, however, human single-unit recordings in MTL revealed also persistently activated neurons supporting WM maintenance (Kamiński et al., 2017; Kornblith et al., 2017; Boran et al., 2019), and amnesia following MTL damage could repeatedly be associated with profoundly compromised WM functionality (Olson et al., 2006; Konkel et al., 2008; Squire, 2017). All these findings argue for a prominent role of MTL in both WM and LTM processing (Jeneson and Squire, 2012; Beukers et al., 2021; Foster et al., in press).

Furthermore, substantial advances in machine-learning analyses (see King and Dehaene, 2014, for an overview) uncovered unexpected coding mechanisms at a neuronal population level to which traditional univariate analyses are blind. One important puzzle piece was the finding that single cells do not necessarily need to be persistently activated for successful online maintenance (Stokes et al., 2013; Sreeni-

vasan et al., 2014). Instead, information may also be mediated via hidden neuronal states as patterns of rapidly changing synaptic weights (Mongillo et al., 2008), which is metabolically less expensive than persistent spiking and may also make memories more robust to interference (Miller et al., 2018). Interestingly, such *'activity-silent'* WM representations can be reactivated, and these reactivations seem to depend on attention and current behavioural relevance (Wolff et al., 2015, 2017; Rose et al., 2016). In line with this, also recent single-unit recordings in MTL provided evidence for such neuronal reactivations after complete activity silence using a picture comparison WM task (Bausch et al., 2021).

Memory contents, finally, can also be stably represented by varying subpopulations of neurons that contribute to the task-related activity at different time courses, and that may even change their individual tuning preferences (Murray et al., 2017; Spaak et al., 2017). Population analyses and advanced projection techniques revealed that such a *dynamic* neuronal selectivity and recruitment, which would conceal online maintenance at a single-cell level, still allows for perfectly robust representations of information (King and Dehaene, 2014).

1.5.2. Functional Implications

The different neuronal codes by which memoranda are expressed – sustained, activity-silent and dynamic – may reflect different processing stages within the memory systems. The influential *embedded-process model* of working memory by Cowan (1988, 1999) proposes three hierarchically arranged states of WM information, (1) long-term memory, (2) a subset of LTM that is currently activated, and (3) the subset of these activated memories that is currently in the focus of attention and awareness. Each of this states operates within its own processing limits. LTM activation, for example, is time-limited; the focus of attention, in contrast, is restricted to a capacity of around 4 items (Cowan, 2001). Importantly, this model incorporates not only a close collaboration between WM and LTM system, it also emphasizes the link between memory and attention which is assumed to be voluntarily controlled by a central executive system (Baddeley and Hitch, 1974; Baddeley, 2012) (in addition to an involuntary attentional orienting system).

Although attempts to find direct neuronal correlates to components of box-and-arrow models are challenging and should always be treated with caution, the observed differential neuronal mechanisms fit quite nicely into Cowan's (1988; 1999) proposed framework of WM. Beukers et al. (2021), for example, draw parallels between activity-silent and episodic memory (Baddeley, 2000), which is part of the LTM and already known to be involved in rapid learning via short-term synaptic plasticity (Zucker and Regehr, 2002). Similarly, others (Kamiński and Rutishauser, 2020; Foster et al., in press) propose that activity-silent memories reflect an intermediate storage in LTM, complementing the active online-maintenance of information in the focus of

attention via persistently activated cells. Kamiński and Rutishauser (2020), finally, suggest that dynamic activity may constitute a neuronal correlate for the central executive which is thought to control attention and the information flow between different memory buffers.

In recent years, the research on working memory has witnessed substantial advancements. Considering WM as a complex interplay of multiple processes including perception, attention, semantic and episodic LTM memory (Rose, 2020) has led to the emergence of new hypotheses and theories regarding the underlying neuronal mechanisms. Countless complex and often seemingly contradictory patterns of behaviour on WM tasks, however, still wait for an explanation.

1.6. Motivation

Behavioural studies with nonhuman primates, indigenous people as well as Western infants, children and adults indicate that we all share an innate 'number sense' that is deeply rooted in human ancestry. Over the decades, numerous studies outlined a complex network encompassing 'core number areas' in frontal and parietal regions, but also sensory and motor areas as well as 'domain-general' structures (including the medial temporal lobe) hosting the memory systems. Valuable pieces were gathered and puzzled together to unravel the neuronal correlates underlying our mental numerical and mathematical representations.

While single-cell recordings are typically conducted in animals and as such strictly limited to nonsymbolic stimulus formats and the simplest arithmetic rules, the technical limitations of most human recording methods prevent the detailed exploration of our anatomical and functional units of the brain: Functional MRI measures neuronal activity only indirectly via blood flow changes, operating on the scale of seconds and with rather low spatial resolution (a single voxel may encompass over 600,000 neurons in cortex), and ECoG, though offering a high temporal resolution, is still limited to combined synaptic mass signals from hundreds of neurons.

In this thesis, we used the rare opportunity to record the activity of single neurons in the medial temporal lobe of awake, behaving neurosurgical human patients that were implanted with chronic intracranial depth electrodes (Fried et al., 2014) to address the following research questions:

- Do single neurons in the human brain encode nonsymbolic and/or symbolic information? If so, how can their tuning properties be characterized?

- Is the behavioural dichotomy observed for subitizing and estimation also reflected in the neuronal response profiles of cells tuned to small and large numerosities?

- How do single units encode abstract arithmetic rules? What do time-resolved recordings in different MTL subregions reveal about the neuronal coding dynamics underlying working memory?

Answering these questions that have yet been eluded from investigation will deepen our understanding of the neurophysiological realization of humans' extraordinary numerical competences.

2. Main Results

In this thesis, we intended to explore fundamental principles of arithmetic processing in the human brain. To that aim, we asked neurosurgical patients that were implanted with chronic intracerebral depth-electrodes in the medial temporal lobe (MTL) to perform simple behavioural tasks while recording neuronal activity of single neurons from parahippocampal cortex (PHC), entorhinal cortex (EC), hippocampus (HIPP) and amygdala (AMY).

In a first study, henceforth referred to as 'calculation task', subjects were instructed to perform simple addition and subtraction tasks on a computer display. After a short fixation period, stimuli were presented sequentially in the order operand 1 – operator – operand 2, each phase being followed by a brief delay during which the stimulus was removed. Afterwards, subjects indicated the calculation result on a number panel on the screen. Task involvement ensured that numbers and arithmetic rules were consciously processed. Due to the sequential task design we were able to explore number (see section 2.1) and rule representations (see section 2.3) separately, both during stimulus presentation and the subsequent working memory phases (delay), respectively.

2.1. Single Neurons in the Human Brain Encode Numbers

At first, we aimed to explore how numerical quantities are represented by single neurons in the MTL. Numerical values of the first operand ranged from 1 to 5, and were presented in different formats, that is, either as nonsymbolic dot arrays or symbolic Arabic numerals. Both formats were shown in standard and control displays (protocols) to control for nonnumerical visual parameters. Number stimuli of operand 2 ranged from 0 to 5, and were always of the same format and protocol as the operand 1 stimulus.

Number conditions varied randomly from trial to trial which allowed us to investigate neuronal responses separately for each of the formats, but also to compare responses to both formats in an unbiased way. In order to avoid confounds with cognitive factors later in the task, we focused primarily on the presentation of operand 1 and the subsequent delay 1 phase.

2.1.1. Single Neurons Respond to Nonsymbolic Numerosities

When the subjects calculated with nonsymbolic dot arrays, about 16 % of all recorded single neurons changed their firing rates dependent on the number of dots (numerosity), irrespective of stimulus appearance (protocol), during presentation of operand 1 and the following delay phase. The highest fractions of these numerosity-selective neurons were found in PHC (29 %) and HIPP (18 %).

For each of these number units, we calculated a tuning function by averaging the firing rates per number condition and determined the unit's preferred numerical value as the number that elicited the strongest average response. The selective neurons' preference covered the entire range of tested numbers 1 to 5. Population filter functions, obtained by combining the tuning curves of all numerosity-selective units preferring the same numerical value, formed overlapping bell-shaped tuning curves that peaked for the preferred number and showed a progressive drop-off of activity the more the number of dots deviated from the preferred value. This systematic decrease was not observed for random tuning curves and reflects a neural correlate of the numerical distance effect (Moyer and Landauer, 1967; Buckley and Gillman, 1974).

2.1.2. Single-Cell Responses to Symbolic Numerals

When the subjects calculated with symbolic numerals, about 3 % of all recorded neurons responded selectively and exclusively to numerals during operand 1 presentation and the subsequent delay phase. Again, the highest fraction of numeral-selective neurons was found in PHC (6 %).

A closer look at the tuning properties of these numeral-selective units revealed striking differences compared to the encoding of numerosities. As before, the number neurons' preference covered the entire tested range of numerical values, and their population tuning curves formed overlapping filter functions. The decline of activity from preferred to nonpreferred numerals, however, was sharp and brisk, and did not differ from tuning curves obtained for randomly shuffled trial labels. This absence of a neuronal distance effect suggests a higher selectivity and more categorical representation of symbolic numbers.

So far, we have looked at each of the two formats individually. As nonsymbolic and symbolic trials were randomly intermixed, we were also able to investigate an individual neuron's response to both formats. A total of 1 % of all units responded selectively to both presentation formats, which was more than expected by chance. Although these cells tended to prefer the same numerical value, the small sample size did not allow the meaningful statistical evaluation of a potential correlation. Furthermore, the preferred number obtained during nonsymbolic trials was not significantly correlated with the (non-significant) number preference during symbolic trials for all numerosity-selective neurons, and vice versa for all numeral-selective neurons. These findings suggest that nonsymbolic numerosities and abstract, symbolic numerals are encoded by two largely segregated neuronal populations.

2.1.3. Neuronal Population Coding

Compared to single neurons, neuronal populations may contain additional information about task contingencies of single neurons (Yuste, 2015; Saxena and Cunning-

ham, 2019). Thus, we adopted a supervised machine learning approach to explore how numerical information was encoded by the two populations of numerosity- and numeral-selective neurons, respectively.

In a first step, support vector machine (SVM) classifiers were trained to identify the numerical value of operand 1 in different trials of the same format, based on the firing rates of the neuronal population at multiple time points of these trials. It was then tested with novel data (i.e. withheld trials of the same format) from the same population to explore how well the number shown in these test trials could be predicted based on the information extracted from the trials during training. We found that the classifiers' accuracy was significantly above chance level (20 % for five number classes) throughout the operand 1 and delay 1 phase for both nonsymbolic- and symbolic-format trials, respectively, albeit with lower performance for the symbolic numbers.

To have a closer look at the type of errors made by the classifiers, we assembled a confusion matrix which summarizes correct and incorrect predictions, broken down by each class. During the time interval of significant number decoding, classifiers trained on nonsymbolic trials tended to confuse numbers that were closer in numerical space more often than ones that were farther apart, reflecting again the numerical distance effect. In contrast, misclassifications of symbolic numerals varied hardly as a function of numerical distance, indicating a sharper transition from the preferred to nonpreferred numerals.

A multi-dimensional state-space analysis, performed separately for numerosity- and numeral-selective neurons, further confirmed these findings. Neuronal population activity can be represented in an n-dimensional space, where each dimension specifies the activity of a single neuron and each point in the space reflects the firing rates of n recorded neurons at a certain time point of the trial. This results in trajectories that are traversed for different neuronal states (i.e. different numerical values) as they evolve over time. While the absolute positions of the trajectories in the state-space are meaningless, relative spatial distances between corresponding points of different states reflect coding differences. Thus, we calculated the Euclidean distances between all pairs of trajectories at a certain time point of the trial. For the nonsymbolic format, the distances between population trajectories systematically increased with numerical distance, starting shortly after onset of operand 1 and persisting until the end of the delay 1 phase. This indicates that patterns of population activity were more similar, the closer two numerosities were in the numerical continuum. For the symbolic format, trajectory differences were confined to the operand 1 phase, and much less pronounced than for the nonsymbolic format, but likewise tended to increase with numerical distance, which may reflect the remnants of a numerical distance effect.

2.1.4. Number Encoding in Later Task Phases

After analysing neuronal responses to operand 1, we also examined number selectivity during the presentation of operand 2. For the nonsymbolic format, about 8 % of all recorded neurons showed activity that varied exclusively with the numerical value of the second operand, irrespective of the dot array layout. The highest fractions of numerosity-selective units were again found in PHC (20 %) and HIPP (6 %). About half of these neurons responded also selectively to the operand 1 stimulus, with a significant proportion of units even being tuned to the same preferred numerosity. In symbolic trials, only a chance proportion of neurons responded exclusively to the numerical value of operand 2.

During this late trial period, neuronal activity may already strongly be modulated by other cognitive factors, like the maintenance of operand 1 as well as the mathematical rule to be applied, or motor response preparation. Still, this findings confirm robust and stable encoding for nonsymbolic number stimuli.

2.2. Distinct Neuronal Representation of Small and Large Numbers in the Human Medial Temporal Lobe

We have shown that the human MTL contains neurons that are selectively tuned to number stimuli (see section 2.1). As the restricted range of small numerical values (1–5) used in the calculation task prevented a detailed exploration of several key aspects of numerical representations, we designed a follow-up study, henceforth referred to as 'parity judgement task'.

After a short fixation period, a numerosity was flashed, followed by a brief delay during which the stimulus was removed. Afterwards, subjects decided as quickly as possible whether the number had been even or odd by pressing a corresponding arrow key as indicated on the screen. Numerosities ranged from 0 to 9, and again different stimulus protocols were used to control for low-level visual features. Task involvement ensured that the participants actively processed the numerical values.

2.2.1. Behaviour

The subjects' performance was in line with behavioural data primarily from developmental psychology (Kaufman et al., 1949; Mandler and Shebo, 1982), showing the typical dichotomy characteristic for subitizing versus estimation. That is, small numerosities ranging from 1 to 4 were judged rapidly and effortlessly with nearly perfect precision. In contrast, numerosities 5 and higher were judged with increasing error rates and reaction times, arguing against serial, symbolic counting (which would be error-free), but rather being indicative of approximate number estimation. Calculating the discontinuity point at which the slope of the behavioural response functions

changed yielded 3.7 (for error rates) and 3.6 (for reaction times) as the average upper boundary of the subitizing range, further bolstering our observation.

The empty set representing 'zero' elicited distinct behavioural effects, probably due to its special status in number concepts (Nieder, 2016b).

2.2.2. Neuronal Responses and Tuning Characteristics

While performing the parity judgement task, a substantial number of single neurons (15 %) was activated exclusively in response to the numerical value of the sample stimulus, and the preference of these number neurons covered the entire range of tested numerosities 0 to 9. The highest proportions of these numerosity-selective neurons were found in PHC (22 %) and HIPP (15 %).

Calculating tuning functions for all individual number neurons yielded the well-known characteristics, i.e. strongest responses to the respective preferred numerosities and a progressive decrease in firing rates with increasing numerical distance. Interestingly, using standardized activity (i.e. z-score relative to baseline activation during fixation) – as opposed to the min/max-normalization applied to the data reported in section 2.1 – unravelled several distinguishing features in the tuning to small versus large numbers.

First, the tuning functions of neurons preferring small numerosities 0–3 showed a systematic surround suppression below spontaneous activity in response to non-preferred numbers; an effect that was not observed in cells preferring large numerosities 4–9. The sharp cut in tuning-flank suppression between 3 and 4 cannot be explained by resolution issues (e.g. the curve becoming too wide to detect suppression), but rather argues for a physiological effect. Next, we fitted Gauss-functions to the tuning curves, and compared the amplitudes and sigmas (as a quantitative measure for tuning width) derived for each curve. Correlating with the former finding, we observed systematic differences in the amplitudes, i.e. significantly smaller values for cells tuned to small versus large numbers, within the two groups, however, tuning amplitudes were indifferent. Similarly, the sigmas for neurons preferring numbers 0–3 were small and indifferent in value (which cannot be explained by a computational floor effect); around preferred number 4 or 5, a turning point emerged with sigmas increasing monotonically. This dichotomy in the response patterns of neurons across the range of numbers parallels our behavioural findings, i.e. narrower, more selective tuning functions linked with more accurate discrimination of smaller numbers, contrasting systematically increasing tuning widths, error rates and reaction times for larger numbers, as expected for ratio-dependent estimation.

Finally, a representational similarity analysis (RSA) based on the correlation coefficients between all pairs of numbers revealed categorically distinct representations of small versus large numerosities. Surprisingly, the correlation matrix suggested a radically different coding for numerosity 0 which remains to be investigated at another

time. In line with our hypothesis, we observed that neurons tuned to the remaining small numbers (1–3) showed more similar firing rates to other small numbers – as reflected by higher correlation coefficients for number pairs from within the same category and rather low values for across-category number pairs –, and vice versa for larger numbers (4–9). To evaluate which number boundary (i.e. 1|2, 2|3, ..., 8|9) segregated the data best into small versus large number representations, we determined the highest and most significant difference between within- and across-category values, which was found for the boundary 3 versus 4.

2.2.3. Coding Differences at the Population Level

To further explore potential decoding discontinuities at the level of neuronal populations, we trained an SVM classifier to discriminate the ten numerosities, based on the firing rates of all number-selective neurons at multiple time points across trial time. Accuracy was significantly elevated above chance level (10% for ten classes) throughout the sample and delay period. Next, we assembled the confusion matrix summarizing correct and incorrect predictions made by the classifier during the time interval of significant number decoding, to have a closer look at the type of errors. Interestingly, we observed that the classifier predominantly confused numerosities from within the small-number category and from within the large-number category, however, it scarcely misclassified numerosities across the two categories which were segregated best at a boundary between numbers 4 and 5.

Finally, we performed a multi-dimensional state-space analysis that examines neuronal population activity as it evolves over time (see also section 2.1). Reducing the high-dimensional space to the three most informative dimensions allows visualizing the neuronal trajectory that is traversed for each of the ten different numerosities (averaged across trials). Spatial closeness (i.e. small distances) of the trajectories indicates similarity in coding, whereas spatial disparity (i.e. large distances) reflects coding dissimilarity. Visual inspection revealed trajectories that were intermingled during fixation but then diverged during sample and delay period, representing numbers with increasing spatial gaps according to ordinal numerical distances (as observed already in the calculation task); an unproportionally large gap, however, segregated the trajectories between 4 and 5.

To statistically quantify this graphical grouping effect, we performed an unsupervised cluster analysis on the neural state-space considering only the time window of significant number decoding. We first determined the optimal number of clusters using two different measures. Although the one criterion was also defined for clustering solutions containing only one cluster (which would be expected under the hypothesis that there are no coding differences between small and large numbers), both measures indicated *two* classes as the optimal cluster number for the dataset.

Partitioning all trials into two classes, the algorithm detected one cluster consisting of the state spaces for numbers 0–4, and a second cluster that comprised the state-spaces for numbers 5–9.

Taken together, two different population-level analyses – supervised SVM classification and unsupervised clustering in the multi-dimensional state-space – confirmed the categorically different encoding of small versus large numerosities observed on the single-cell level, indicating a boundary between numbers 4 and 5. Notably, when applying z-score normalization to the number data from the calculation task, we observed the same characteristic coding features distinguishing small and large numbers as reported above, both on single-unit- and population-level (unpublished data), strongly corroborating the robustness of our findings.

2.3. Neuronal Codes for Arithmetic Rule Processing in the Human Brain

We have shown that single neurons in the human MTL stably represent numerical cardinality. Calculating with numbers, however, requires not only the online maintenance of number information, but also of the arithmetic rules according to which these numbers are to be manipulated. We therefore investigated whether these single neurons and neuronal populations also encode mentally performed calculations.

In the calculation task, we applied two arithmetic rules, namely addition and subtraction, that were instructed either as arithmetic signs or analogous written words. Using different rule notations (rule cues) allowed us to dissociate neural activity related to low-level visual features of the operator from the abstract rule that it represented.

2.3.1. Single Neurons Respond to Calculation Rules

At first, we identified individual neurons that selectively enhanced their neuronal activity according to the arithmetic rule. During presentation of the operator, a small but significant proportion (about 5 %) of all recorded neurons was modulated by the arithmetic rule, most of these units (3.5 %) being exclusively rule-selective, i.e. showing no effects for the notation of the rule cue or any other task-relevant factor. A closer anatomical look showed fundamental differences in the unit proportions across different subregions of the MTL. While EC and AMY amygdala exhibited rule-selective neurons only as expected by chance, significant fractions of rule units were observed in PHC (7 %) and HIPP (4 %). In PHC, in particular, a relatively large proportion of neurons was also responsive to the rule cue.

In the subsequent rule delay period, this proportion increased to 5.3 %, significant fractions now being observed in all subregions. Notably, a significant proportion of HIPP units responded also to the number and format of operand 1 during this phase,

probably reflecting active maintenance of information about both the first operand and the arithmetic rule in working memory.

With trial progression, the proportion of rule-selective units dropped to chance level. Instead, cells encoded the additionally introduced task factors. Both in PHC and HIPP we found units that discharged selectively for the second operand during the operand 2 and following delay 2 period, but also for the calculation result during the delay 2 period, seemingly multiplexing all information necessary for the calculation.

2.3.2. Notation-Independent Representation of Addition and Subtraction Rules

Though undoubtedly the ineluctable first analysis step, focussing only on single cells may underscore the importance of complex spatio-temporal patterns of population activity by which information can also be represented. Using a statistical classification approach, we therefore explored how the entire population of recorded units encoded rule information. In order to identify potential regional differences, the four subregions of the MTL were analysed separately. For that, we trained SVM classifiers to discriminate between addition and subtraction trials, combining the respective cues per rule, based on the firing rates at different time points across trial time, and then tested the models on novel data from the same population.

Information about the arithmetic rule was successfully decoded in all MTL areas during long time intervals of the trial period. The response patterns, however, varied considerably between the four areas. Consistent with the single-cell analysis, effects were strong and long-lasting in PHC and even more pronounced in HIPP. In the PHC, accuracy ramped up shortly after presentation of the operator and again after presentation of operand 2, but returned to chance level in between and shortly after operand 2 offset. In HIPP, classification accuracy peaked after cue presentation, remained stable throughout the operand 2 phase, and ramped even further up until it reached its maximum during delay 2, i.e. when all information was available to perform the calculation. In EC and AMY, in contrast, significant effects were rather weak and short-lived, being confined primarily to the delay 2 phase.

High classification accuracies do not imply *per se* that a classifier has learned to encode 'abstract' rule information; comparable values might also be observed if the classifier had learned to encode one rule cue perfectly, but remained at chance level for the other three cues. To account for this, we looked at the classification probabilities per rule cue during the previously found significant time intervals. In all subregions, classifier performance was significant for addition and subtraction across both rule cue notations. Moreover, classification probabilities were comparable for all cues (except in AMY), confirming the abstractness of the encoded information.

The population analyses do not yet rule out that neurons may encode the rule cue in addition to the rule itself. To control for this, we also investigated how cue information was encoded, training classifiers to discriminate between the different

cue notations (sign or verbal analogue) while combining both arithmetic rules. Reliable and significant decoding of the cue information was possible only in PHC, and exclusively during the presentation of the operator. This result is in line with the effects observed in the single-cell analysis, and further corroborates our finding that the population response patterns recorded during later task phases indeed reflect abstract rule presentations.

2.3.3. Cross-Notation Decoding of Addition and Subtraction

The previous classification analyses indicated that arithmetic rules were encoded irrespective of the specific rule cue. To put this observation to the test, we finally analysed the neuronal populations' ability to generalize rule information across different cue notations. For that, we performed another classification analysis, training a classifier on trials of one rule cue, and testing the model on trials of the other rule cue, and vice versa. Generalization was then judged successful, if (1) synchronous intervals of significant classification were found for both directions of generalization, and if (2) the accuracies averaged across generalization directions were significant for each arithmetic rule in these synchronous time windows.

In PHC and HIPP, we observed significant cross-notation decoding in extended and overlapping time windows. The temporal performance profiles were very similar for both directions of generalization, and strongly resembled the patterns observed for the full dataset. That is, two selective periods interrupted by a nonselective period in PHC, and a prolonged period of significant decoding in HIPP, emerging after cue offset and up to the end of the trial. In both areas, the accuracy of transfer was significant for both arithmetic rules and both generalization directions. EC and AMY, in contrast, failed to generalize across different rule cue notations. In EC, classification accuracy reached significance briefly, but for only one test direction. In AMY, decoding performance was at chance level throughout the whole trial for both generalization directions. Thus, both areas failed our criterion for successful cross-notation decoding.

2.3.4. Cross-Temporal Calculation Rule Decoding

We have shown that populations of neurons in the MTL represent information about simple arithmetic rules. More importantly, intervals of significant decoding were observed not only during presentation of the operator, but also in the subsequent delay and later task phases. This indicates that rule information was maintained actively in working memory until all information necessary for calculation was available. To investigate the underlying neuronal codes of arithmetic rules in more detail, we next performed a temporal cross-training analysis.

Traditional sliding-window decoding approaches train and test statistical classifiers on data from identical time points across trial time. The temporal generalization method (King and Dehaene, 2014) extends this approach by testing each model on *all* time points, resulting in a temporal generalization matrix in which each row corresponds to the time at which the decoder was trained and each column to the time at which it was tested. Distinct decoding patterns observed in the temporal generalization matrix then allow to characterize the time course of population codes in more detail, and to draw conclusions about differences in the underlying neuronal codes for working memory. Persistently activated neurons manifest in a *static population activity* that is also stable over time. In other words, a classifier trained on time t_1 would still be able to decode the information when tested on time t_2, resulting in a square-shaped decoding profile in the matrix. Alternatively, assemblies of sequentially activated, sparsely firing neurons with rapidly changing tuning profiles may maintain memory contents via *dynamic population codes* that would allow no or only little cross-temporal generalization, as reflected by a high decoding performance along the main diagonal of the matrix and a strong reduction of the off-diagonal values.

In PHC, we observed high classification accuracies only along the main diagonal of the temporal generalization matrix. That is, a model trained on the firing rates observed, for example, during cue presentation, was able to decode rule information only during that time window, but failed to do so when tested on activity recorded during presentation of operand 2 or other time points. Or in other words, rule decoding was only successful if training and test time of the classifier were identical. Notably, we could also identify this coding profile when testing for generalization across cue notations (as reported above), even if only to a weaker extent. This absence of cross-temporal generalization indicates that PHC neurons rapidly change their tuning properties with time, resulting in a dynamic neuronal population code.

A rather different picture emerged for the HIPP. We observed stable significant cross-temporal generalization starting at the end of the rule cue period until the end of the trial. To put it another way, a classifier trained on firing rates recorded, for example, during cue presentation, was still able to decode the rule information when tested on activity recorded during presentation of operand 2 or even later. Again, this temporal coding profile was still observed when testing for generalization across cue notations. This square-like accuracy pattern argues for a rather static neuronal population code, probably based on cells that are persistently active across trial time.

Finally, EC showed a mild square-like accuracy pattern that emerged around the rule delay, suggesting rather stable rule coding that vanished with presentation of operand 2. The weak effects observed in the AMY did not allow any statements about the underlying coding dynamics.

3. Discussion

Decades of research on numerical and mathematical cognition have outlined a complex maths network in the human brain, comprising 'core number areas' in frontal and parietal regions (Arsalidou and Taylor, 2011), but also sensory and motor areas as well as more 'domain-general' areas (including the medial temporal lobe) hosting, for example, the memory systems (Menon, 2016).

In numerous studies with nonhuman primates, number- and rule-selective neurons with distinctive tuning characteristics were found in these number areas (Nieder, 2016a). Although directly linking neuronal responses with behavioural judgements, these findings are strictly limited to nonsymbolic stimuli and the simplest arithmetic rules. In contrast, although studies with humans suggest anatomical and functional homologues, recording methods like fMRI or ECoG do not allow the detailed exploration of single neurons, leaving many important questions unanswered.

In this thesis, we bridged this gap by recording activity of single cells in the MTL of human subjects, addressing aspects that have yet been eluded from investigation.

3.1. Number Neurons in the Human Brain

In a first study, we asked subjects to perform simple calculation tasks, aiming to identify and characterize neurons responsive to numerical information. Displaying operands and operator sequentially allowed us to explore pure number representations, detached from confounds with other task-related factors. Furthermore, operands were presented either as nonsymbolic dot arrays or symbolic Arabic numerals. Randomly varying the format from trial to trial, we were able to analyse neuronal responses to each of the formats individually. Thus, we can now compare data about nonsymbolic number coding in humans to those of nonhuman primates. Likewise, we can investigate how symbolic number is represented in this part of the human brain, addressing also the still unresolved question whether neuronal responses are abstracted beyond presentation formats.

The restricted range of numerical values (1–5) used in the calculation task prevented the detailed exploration of several key aspects of numerical representations. Specifically, we aimed to analyse whether the behavioural dichotomy observed for subitizing and estimation is also reflected in the neuronal response profiles of single cells tuned to small and large numerosities. Thus, we designed a follow-up study using a simple parity judgement task that required the subjects to indicate whether nonsymbolic dot arrays were even or odd. In this second task, we used numerical values 0–9 to cover both the subitizing and the estimation range.

3.1.1. Encoding of Numerical Information by MTL Single Neurons

As the crucial first step, we showed that a substantial proportion of cells in the human MTL responds selectively to the numerical values of number stimuli, thereby finally providing the direct human homologue to monkeys' number-selective neurons that had so far only been hypothesized based on functional and (large-scale) anatomical parallels in human imaging studies.

Among the four MTL regions we recorded from, the PHC showed the highest proportions of number units, followed by the HIPP. The large fractions of number neurons we found in the calculation task were comparable to the proportions observed during the parity judgement task. This concurs also with numerical tuning of HIPP neurons recently observed in nonhuman primates (Opris et al., 2015). These consistent findings in different subject cohorts and with different task protocols indicate that PHC and HIPP contribute significantly to numerical representations.

The PHC is highly interconnected with other polymodal association areas, including the parieto-frontal number system (Goldman-Rakic et al., 1984; Suzuki, 2009). As such, representations about numerical magnitudes do most likely not originate within the PHC (or other areas of the MTL), instead, semantic information about numbers may rather be provided via direct connections from parietal and frontal core number areas.

The MTL is a highly associative brain area contributing to many cognitive processes (Aminoff et al., 2013). As part of the declarative memory system, it hosts 'concept cells' (Quian Quiroga, 2012) characterized by remarkable selectivities to particular categories. Considering the enormous importance of quantity perception in cognition, our finding of a neuronal substrate for the category 'numerosity' may thus not be completely unexpected. Furthermore, based on observed responses to spatial factors (Ekstrom et al., 2003; Jacobs et al., 2013) and mirror actions (Mukamel et al., 2010), it has been speculated that the MTL may also play a role in a sensorimotor numerosity system that links action to magnitude perception (Anobile et al., 2021). Such speculations as well as other functional implications need yet to be explored.

3.1.2. Segregated Populations of Numerosity- and Numeral-Selective Cells

Our analyses unveiled two largely segregated populations of tuned number neurons that process either nonsymbolic or symbolic quantity; abstract cells that encoded the same numerical value in both formats were rarely found.

For many years, there have been heated discussions in the human functional imaging literature whether neuronal representations of number in the IPS (one of the most important number areas) were format-independent (Piazza et al., 2007; Eger et al., 2009; Damarla et al., 2016) or -dependent (Cohen Kadosh and Walsh, 2009). Thankfully, in the last years research focus has finally moved away from the dichotomous

'abstract or not?'. Admitting coding differences without ruling out the possibility of abstract number processings, it instead shifted more to the conciliating questions of where and how abstract and nonabstract number representations coexist. Despite the progress that has been made, however, the technical limitations of fMRI and ECoG prevent answering important questions regarding the degree of abstractness. Most critically, whether observed differential activations for nonsymbolic and symbolic stimuli stem from different subpopulations of format-sensitive single neurons with distinctive tuning properties, or rather from the same population of neurons that are differentially modulated by different formats remains eluded at the macroscopic voxel scale of these recording techniques.

Overcoming this limitation, our single-unit recordings argue for the former possibility (at least for cells in the MTL) of distinct populations of tuned neurons that represent either nonsymbolic or symbolic numerical information. Future single-cell recordings in human subjects, in particular in the parietal and frontal association cortices, may help further resolve the question of abstract or segregated number neurons.

3.1.3. Neuronal Codes for Numbers

Irrespective of its neurophysiological realization, format dependency does not pose a conceptual problem to number coding. Two competing hypotheses have been proposed: Numbers could either be encoded by a 'summation code', characterized by monotonic activations that vary as a function of quantity (Roitman et al., 2007), or by a 'labelled-line code' as witnessed by numerosity-selective units tuned to preferred numerosities (Nieder and Merten, 2007). Influential computational models of number processing (Dehaene and Changeux, 1993; Verguts and Fias, 2004) showed that the two codes are not mutually exclusive, but suggested summation coding only for an intermediate processing stage.

For both numerosity- and numeral-selective neurons, we observed activations that peaked for a preferred quantity and were systematically modulated by numerical distances. Forming overlapping tuning functions inherently ordered by ordinal numerosity, these cells covered the whole investigated number space unintermittedly. These findings show striking similarities with the tuning profiles found multiple times in single-cell recordings of monkeys, both in trained (Nieder et al., 2002, 2006; Sawamura et al., 2002; Nieder and Miller, 2004; Nieder, 2012) and numerically naive animals (Viswanathan and Nieder, 2013), and even in corvid birds (Ditz and Nieder, 2015). This indicates that – at this advanced level within the cognitive processing pipeline – number coding in the brain of humans and other animals is best captured by a labelled-line code.

3.1.4. Distinct Representations of Nonsymbolic and Symbolic Quantities

Interestingly, we observed also striking differences in the tuning of numerosity- and numeral-selective neurons. One of the most important metrics for indexing neuronal number representations is proving that neurophysiological activity is parametrically modulated by the numerical distance effect. Indeed, in both tasks, the neuronal representations of nonsymbolic numerosities were abundant and showed a marked distance effect. In contrast, representations of symbolic numerals were sparsely found, and their tuning curves were rather brisk and categorical. Population analyses further corroborated this findings. In the neuronal state-space, inter-trajectory distances increased systematically with increasing numerical distance. These effects were very strong and long-lasting for numerosity-selective neurons, but rather short-living and less pronounced for numeral-selective cells, probably reflecting remnants of a distance effect.

In order to link number neurons to numerical behaviour, neuronal responses need to explain number judgements (Nieder and Miller, 2003; Pinel et al., 2004). Of course, providing a direct correlate would have been very informative, unfortunately, the calculation task did not ask for an explicit behavioural response to the single operands, and our participants hardly made any mistakes precluding also the evaluation of error trials (an analysis regularly done in monkeys). Our findings are in agreement with behavioural studies, though, that report a distance effect that is strongly pronounced for nonsymbolic stimuli, but minute for judgements of exact number symbols (Buckley and Gillman, 1974). Furthermore, we showed that numerical information was robustly decoded by statistical classifiers from all neurons tuned to numerosities, and with lower accuracy also from the population of numeral-selective neurons. Misclassifications followed the same distinct patterns as observed for the tuning curves of individual cells. This held true also for the entire population of recorded neurons and irrespective of response selectivity, thus fulfilling a basic requirement to link neurons and behaviour (Ramirez-Cardenas et al., 2016).

The capacity to represent nonsymbolic quantities traces all the way back to our monkey ancestors who had evolved a 'number sense' that favoured speed over precision to gain evolutionary advantages over competitors. The highly precise symbolic enumeration system, in contrast, is something uniquely human (Nieder, 2009) and may as such be a special feature of the human brain. Indeed, some researchers challenge a neurobiological link between number sense and symbolic numerical abilities (Wilkey and Ansari, 2019). There is, however, also evidence that the ANS plays an important role in the cognitive development of symbolic numerical thinking (Halberda et al., 2008; Piazza, 2010; Szkudlarek and Brannon, 2017), and the capacity to link number to arbitrary shapes has also precursors in nonhuman primates (Diester and Nieder, 2007, 2010; Livingstone et al., 2014). The finding of a distance effect for both formats, together with the observed differences in coding precision – that are

indeed in line with both human fMRI studies (e.g. Piazza et al., 2007; Eger et al., 2009; Lyons et al., 2015) and computational models (Verguts and Fias, 2004) – thus argue for the hypothesis that precise high-level numerical abilities are grounded in neuronal circuits devoted to deriving precise numerical values from the evolutionary older approximate numerosity representations (Dehaene and Cohen, 2007).

3.1.5. Coding Dichotomy for Small and Large Numerosities

In the parity judgement task, behavioural measures showed fast and error-less responses for small numbers up to four, and increasingly slower and error-prone responses for larger numerosities. The findings are perfectly in line with the well-known dichotomous effects characteristic for subitizing versus estimation (Kaufman et al., 1949; Mandler and Shebo, 1982).

Interestingly, we observed several distinguishing features in the tuning to small versus large numerosities of single units that mirrored this behavioural dichotomy. While tuning width for large numbers increased in a ratio-dependent manner indicative of estimation, neuronal tuning to small numbers was more selective and ratio-independent. This coding dichotomy was also confirmed at the population level. The boundary in neuronal coding around number the 'magical number 4' (Cowan, 2001), consistently observed across all analyses, correlated well with the behavioural transition from subitizing to estimation. This indicates that numbers in the subitizing range may be tapped by different mechanisms in addition to that for number estimation (Anobile et al., 2016).

Furthermore, neuronal tuning in the subitizing range was characterized by a distinct surround suppression below baseline activity for nonpreferred numerosities that was not observed for large numbers. Lateral inhibition is a basic neuronal circuit operation (Hartline et al., 1956), generated by excitatory neurons firing in response to their preferred stimulus and concomitantly recruiting broadly-tuned inhibitory interneurons which, in turn, suppress the firing of neurons tuned to different preferred stimuli. It is known to increase contrast sensitivity and to shape the tuning of cortical neurons (Isaacson and Scanziani, 2011), and could thus mechanistically explain the more accurate number discrimination we observed in the subitizing range. Indeed, it has been shown that inhibition via interneurons sharpened the tuning to numerosities in the animal brain (Diester and Nieder, 2008; Ditz et al., 2022). Similarly, several computational models suggest that centre-surround selectivity profiles emerge spontaneously in unsupervised neural networks (Stoianov and Zorzi, 2012; Nasr et al., 2019), and that task-dependent differences in the inhibition strength do not only give rise to capacity limitations but can also explain the activation differences between subitizing and estimation range (Sengupta et al., 2014; Knops et al., 2014).

Extensive research on the subitizing phenomenon suggests a pivotal role of attention and WM processes that enhance rather than replace estimation for small numbers

(Anobile et al., 2016). In their attempts to approach neuronal correlates, however, human functional imaging studies are confined to a rather macroscopic level of spatial resolution. To our knowledge, we are the first to provide a 'cellular footprint' for subitizing directly on a single-cell level. Whether the observed tuning-flank suppression originates within the MTL or is rather transferred from other areas involved in attention- or WM-related processes remains to be investigated. With excitatory and inhibitory neurons identified in the human MTL (Ison et al., 2011; Gast et al., 2016; Mosher et al., 2020), though, the necessary circuit components would readily be available to implement surround inhibition for selective coding in the subitizing range.

Future research that complements the parity judgement task with richer and more explicit number tasks, and that directly contrast the responses of neurons with and without attentional demands assigned to number representation will help to support the generality of our findings.

3.2. Neuronal Codes for Abstract Arithmetic Rules

In the first experiment, the calculation task, we asked subjects to perform simple addition and subtraction tasks. The sequential presentation of operands and operator allowed us not only to analyse number representations, but also to explore pure rule processing, reducing confounds with number representations of the operands and other task-related factors. To control also for low-level visual features, arithmetic rules were indicated by cues in two different notations (arithmetic sign or verbal analogue).

3.2.1. Single Neurons Selective for Numerical Rules

As a first step, we identified single neurons encoding the arithmetic rule, irrespective of the concrete cues indicating that rule. A significant proportion of these exclusively rule-selective single neurons was detected right after the presentation of the operator. This rule-selectivity diminished with the presentation of the second operand and the ongoing calculation in favour of other task factors.

So far, neuronal correlates of addition and subtraction have not been studied in monkeys. Selective responses of single neurons to abstract 'greater/less than' rules, though, have been found in parietal and frontal areas of rhesus monkeys (Bongard and Nieder, 2010; Eiselt and Nieder, 2013). The PFC, in particular, is associated with the representation of abstract rules and concepts (Mansouri et al., 2020) within the neuronal 'core maths network'. The MTL, however, may also have direct access to calculation-relevant information, as it is highly interconnected with these neocortical association areas (Goldman-Rakic et al., 1984; Suzuki, 2009). Indeed, the presence of number-selective neurons that we observed in our study, and that was also reported for monkey HIPP neurons (Opris et al., 2015), indicates that the MTL is suited to

manipulate number representations. As such, it could mediate the transformation of perceived numerical information in a working memory buffer.

The rule-selectivity reported in nonhuman primates showed a substantial degree of specialization, responding only to rules applied to a specific magnitude type (Eiselt and Nieder, 2013). We therefore think that the majority of our rule neurons specifically and genuinely encode arithmetic rules. However, this does not rule out the possibility that some of these neurons may not also become engaged in the encoding of other rules (e.g. the 'even/odd' rule in the parity judgement task) that we have not yet explored. Indeed, response-selectivity is not a fixed, inalterable feature of single neurons. It has been shown that learning and memory training can alter neuronal selectivity (Qi et al., 2011). Similarly, responses to the same stimuli may differ as a function of task demands (White and Wise, 1999; Asaad et al., 2000; Viswanathan and Nieder, 2015). Such dynamic changes in coding capacities were, for example, also reflected in reduced proportions of number-selective cells in monkeys performing a rule-based numerical discrimination task (Vallentin et al., 2012; Eiselt and Nieder, 2013), compared to the amount of number neurons recorded from the same animals performing a simple delayed match-to-numerosity task (Nieder et al., 2002; Nieder and Merten, 2007; Diester and Nieder, 2008). In the same vein, the top-down interplay with the PFC may also lead to a flexible numerical coding in the MTL, probably explaining the diminished rule-selectivity we observed with the presentation of the second operand and the ongoing calculation.

3.2.2. Online Maintenance of Rule Information in Working Memory

Classification analyses revealed that the entire population of recorded neurons carried sufficient information about the arithmetic rules to discriminate between addition and subtraction during mental calculation. Importantly, decoding generalized across different rule cue notations which implies an abstract and notation-independent representation of these rules.

When people solve simple arithmetic problems, they can either retrieve rote numerical facts stored in declarative memory, or they have to apply alternative strategies usually involving WM processes. The restricted numerical range used in our task, which permitted only calculation results between 0 and 9, makes it likely that subjects retrieved their responses primarily from rote memory. The sequential stimulus presentation, however, demanded the parallel maintenance of the individual task components until all information necessary for solving the task was available, thus recruiting also WM resources. Indeed, in PHC and HIPP, successful decoding of rule information was observed from the beginning of the rule cue presentation until the end of the trial, indicating that the neuronal responses were detached from the physical stimulus, instead reflecting active maintenance in WM. This finding is in line with previous intracranial recordings in humans that reported persistent activations

of MTL single neurons for the same preferred stimuli throughout several seconds of temporal gaps, showing that this delay activity correlated with memory load and predicted the successful retrieval of WM contents (see Rutishauser et al., 2021, for a review).

3.2.3. Static and Dynamic Codes for Working Memory

Decoders applied to time-resolved recordings unveiled fundamentally different coding profiles of WM in different subregions of the MTL. In HIPP, we observed robust across-time generalization during mental calculation characteristic for static neuronal coding (King and Dehaene, 2014). This coding pattern may originate from persistently activated MTL neurons, as reported also by other human single-unit studies (Kamiński et al., 2017; Kornblith et al., 2017; Boran et al., 2019). In PHC, in contrast, decoders did not generalize across different time points, clearly indicating a dynamic coding framework (King and Dehaene, 2014). Such a decoding pattern has also been observed in monkeys during complex WM tasks, and may presumably be explained by varying subpopulations of neurons with rapidly changing tuning preferences (Murray et al., 2017; Spaak et al., 2017).

Static and dynamic codes are not incompatible, but may instead hint at distinct cognitive functions in arithmetic for the two MTL subregions. A recent approach by Kamiński and Rutishauser (2020) proposes that differential neuronal coding mechanisms can be associated with different components of Cowan's embedded-process model of WM (Cowan, 1988, 1999). Persistent activity is assumed to reflect activated long-term memories in the current focus of attention; dynamic activity, in contrast, indicates active processing via the central executive that is thought to implement changes in the focus of attention. Following this logic, the dynamic coding patterns we observed in PHC may reflect attentional shifts to short-term representations of the arithmetic rule, whereas downstream HIPP may 'do the maths' based on the attended memory contents activated from LTM in order to manipulate the operands according to the arithmetic rule at hand. Of course, many questions are still unanswered and more fine-grained analyses are required to decipher the individual roles of different brain areas and neuronal mechanisms in mental arithmetic.

3.3. Conclusion

Using the rare opportunity to record the activity of single neurons in the medial temporal lobe of behaving humans, we were able to bridge the gap between single-unit recordings in animals (that inherently prevent exploring the full spectrum of humans' cognitive abilities), and functional imaging studies in humans (whose technical limitations confine probing cognition to a rather macroscopic level). Like a prism that breaks light up into all colours of the rainbow, two different numerical tasks investigated in

this thesis shed light on seemingly disparate aspects of numerical and mathematical cognition:

- We proved that single cells in the MTL can encode information about both numerical quantities and simple arithmetic rules, thereby finally providing the direct human homologue to monkeys' number- and rule-selective neurons that had so far only been hypothesized based on functional and (large-scale) anatomical parallels in human imaging studies.
- Numerical representations follow a labelled-line coding as observed also in animals. The finding of segregated populations of numerosity- and numeral-selective neurons that respond to different stimulus formats with distinct tuning profiles, however, indicates different degrees of abstractness by which quantities are encoded in the human number network.
- The observed coding dichotomy for small and large numerosities, mirroring subitizing and estimation processes, provides a first 'cellular footprint' on that topic which may help to better understand the neuronal computations underlying the complex interplay of attention, working memory, and number representations.
- Finally, revealing static and dynamic coding mechanisms in PHC and HIPP does not only emphasize the MTL's role as an integral part of a wider cortical maths network, but more importantly, it highlights the substantial role it also plays in WM processes.

As always in science, the new discoveries raise new questions. For instance, whether the current findings in the MTL transfer to other cortical brain regions. Taken together, this thesis provides valuable puzzle pieces that deepen our understanding of numerical representations constituting the 'sense of number'.

Abbreviations

aITC	anterior inferior temporal cortex
AMY	amygdala
ANS	analogue number system
DCE	direct cortical electrostimulation
dlPFC	dorsolateral prefrontal cortex
DoG	difference-of-Gaussian
EC	entorhinal cortex
ECoG	electrocorticography
EEG	electroencephalography
ERP	event-related potentials
fMRI	functional magnetic resonance imaging
HIPP	hippocampus
IPS	intraparietal sulcus
LTM	long-term memory
MTL	medial temporal lobe
OTS	object tracking system
PET	positron emission tomography
PFC	prefrontal cortex
PHC	parahippocampal cortex
PPC	posterior parietal cortex
RSA	representational similarity analysis
SVM	support vector machine
WM	working memory

References

Agostino A, Johnson J, Pascual-Leone J (2010) Executive functions underlying multiplicative reasoning: Problem type matters. *Journal of Experimental Child Psychology* 105:286–305.

Allison T, Puce A, Spencer DD, McCarthy G (1999) Electrophysiological studies of human face perception: I. potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex* 9:415—-430.

Aminoff EM, Kestutis K, Bar M (2013) The role of the parahippocampal cortex in cognition. *Trends in Cognitive Sciences* 17:379–390.

Anderson JS, Ferguson MA, Lopez-Larson M, Yurgelun-Todd D (2010) Topographic maps of multisensory attention. *Biological Sciences* 107:20110–20114.

Anderson JR, Schooler LJ (1991) Reflections of the environment in memory. *Psychological Sciences* 2:396–408.

Anobile G, Cicchini G, Burr D (2016) Number as a primary perceptual attribute: A review. *Perception* 45:5–31.

Anobile G, Arrighi R, Burr DC (2019) Simultaneous and sequential subitizing are separate systems, and neither predicts math abilities. *Journal of Experimental Child Psychology* 178:86–103.

Anobile G, Arrighi R, Castaldi E, Burr DC (2021) A sensorimotor numerosity system. *Trends in Cognitive Sciences* 25:24–36.

Ansari D (2007) Does the parietal cortex distinguish between '10', 'ten' and ten dots? *Neuron* 53:165–167.

Ansari D (2008) Effects of development and enculturation on number representation in the brain. *Nature Reviews Neuroscience* 9:278–291.

Arrighi R, Togoli I, Burr DC (2014) A generalized sense of number. *Proceedings of the Royal Society B* 281:20141791.

Arsalidou M, Pawliw-Levac M, Sadeghi M, Pascual-Leone J (2018) Brain areas associated with numbers and calculations in children: Meta-analyses of fMRI studies. *Developmental Cognitive Neuroscience* 30:239–250.

Arsalidou M, Taylor MJ (2011) Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. *NeuroImage* 54:2382–2393.

Asaad WF, Rainer G, Miller EK (2000) Task-specific neural activity in the primate prefrontal cortex. *Journal of Neurophysiology* 84:451–459.

Ashcraft MH (1992) Cognitive arithmetic: A review of data and theory. *Cognition* 44:75–106.

Ashcraft MH (1995) Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition* 1:3–34.

Atkinson R, Shiffrin R (1968) Human memory: A proposed system and its control processes In Spence K, Spence J, editors, *The Psychology of Learning and Motivation: Advances in Research and Theory*, pp. 89–195. New York: Academic Press.

Baddeley A, Hitch G (1974) Working memory In Bower G, editor, *The Psychology of Learning and Motivation*. New York: Academic Press.

Baddeley A (2000) The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4:417—-423.

Baddeley A (2003) Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* 4:829–839.

Baddeley A (2012) Working memory: Theories, models, and controversies. *Annual Review of Psychology* 63:1–29.

Bausch M, Niediek J, Reber TP, Mackay S, Boström J, Elger CE, Mormann F (2021) Concept neurons in the human medial temporal lobe flexibly represent abstract relations between concepts. *Nature Communications* 12:6164.

Beukers A, Buschman T, Cohen J, Norman K (2021) Is activity silent working memory simply episodic memory? *Trends in Cognitive Sciences* 25:284–293.

Bongard S, Nieder A (2010) Basic mathematical rules are encoded by primate prefrontal cortex neurons. *Proceedings of the National Academy of Sciences, USA* 107:2277–2282.

Booth J, Siegler R (2006) Developmental and individual differences in pure numerical estimation. *Developmental Psychology* 41:189–201.

Boran E, Fedele T, Klaver P, Hilfiker P, Stieglitz L, Grunwald T, Sarnthein J (2019) Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. *Science Advances* 5:eaav3687.

Boyer CB (1944) Zero: The symbol, the concept, the number. *National Mathematics Magazine* 18:323–330.

Boyer CB (1968) *A history of mathematics* New York & London: John Wiley & Sons.

Brannon E, Terrace H (1998) Ordering of the numerosities 1 to 9 by monkeys. *Science* 282:746–749.

Buckley PB, Gillman CB (1974) Comparisons of digits and dot patterns. *Journal of Experimental Psychology* 103:1131–1136.

Burr D, Ross J (2008) A visual sense of number. *Current Biology* 18:425–428.

Burr DC, Turi M, Anobile G (2010) Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision* 10:1–10.

Cai Y, Hofstetter S, van Dijk J, Zuiderbaan W, van der Zwaag W, Harvey B, Dumoulin S (2021) Topographic numerosity maps cover subitizing and estimation ranges. *Nature Communications* 12:3374.

Camos V, Tillmann B (2008) Discontinuity in the enumeration of sequentially presented auditory and visual stimuli. *Cognition* 107:1135–1143.

Campbell JI (2008) Subtraction by addition. *Memory & Cognition* 36:1094–1102.

Campbell JI, Timm J (2000) Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review* 7:692–699.

Campbell JI, Xue Q (2001) Cognitive arithmetic across cultures. *Journal of Experimental Psychology* 130:299–315.

Cantlon JF, Brannon EM (2006) Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science* 17:401–406.

Carey S (2009) *The Origin of Concepts* Oxford: Oxford University Press.

Carey S, Barner D (2019) Ontogenetic origins of human integer representations. *Trends in Cognitive Sciences* 23:823–835.

Carey S (2002) Cognitive foundations of arithmetic: Evolution and ontogenesis. *Mind & Language* 16:37–55.

Chen BL, Hall DH, Chklovskii DB (2006) Wiring optimization can relate neuronal structure and function. *Proceedings of the National Academy of Sciences, USA* 103:4723–4728.

Cheyette SJ, Piantadosi ST (2020) A unified account of numerosity perception. *Nature Human Behavior* 4:1265–1272.

Cho S, Metcalfe A, Young C, Ryali S, Geary D, Menon V (2012) Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. *Journal of Cognitive Neuroscience* 24:1849–1866.

Christoff K, Gabrieli JD (2000) The frontopolar cortex and human cognition: Evidence for a rostrocaudal hierarchical organization within the human prefrontal cortex. *Psychobiology* 28:168–186.

Cicchini G, Anobile G, Burr D (2016) Spontaneous perception of numerosity in humans. *Nature Communications* 7:12536.

Cohen Kadosh R, Cohen Kadosh K, Kaas A, Henik A, Goebel R (2007) Notation-dependent and -independent representations of numbers in the parietal lobes. *Neuron* 53:307–314.

Cohen Kadosh R, Walsh V (2009) Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences* 32:313–328.

Cowan N (1988) Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin* 104:163–191.

Cowan N (1999) An embedded-process model of working memory In Miyake A, Shah P, editors, *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, pp. 62–101. Cambridge: Cambridge University Press.

Cowan N (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral Brain Sciences* 24:87–114.

Curtis C, D'Esposito M (2004) The effects of prefrontal lesions on working memory performance and theory. *Cognitive, Affective & Behavioral Neuroscience* 4:528–539.

Daitch AL, Foster BL, 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, USA* 113:E7277–E7286.

Damarla SR, Cherkassky VL, Just MA (2016) Modality-independent representations of small quantities based on brain activation patterns. *Human Brain Mapping* 37:1296–1307.

Dayan P, Abbott LF (2005) *Theoretical neuroscience: Computational and mathematical modeling of neural systems* MIT Press.

De Smedt B, Holloway I, Ansari D (2011) Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetic fluency. *NeuroImage* 57:771–781.

Dehaene S, Bossini S, Giraux P (1993) The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General* 122:371–396.

Dehaene S, Changeux J (1993) Development of elementary numerical abilities: A neuronal model. *Journal of Cognitive Neurosciences* 5:390–407.

Dehaene S, Dehaene-Lambertz G, Cohen L (1998) Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience* 21:355–361.

Dehaene S, Izard V, Spelke E, Pica P (2008) Log or linear? distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science* 320:1217–1220.

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

Dehaene S (1996) The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience* 8:47–68.

Dehaene S (1997) *The number sense: How the mind creates mathematics* Oxford University Press.

Dehaene S (2007) Symbols and quantities in parietal cortex: Elements of a mathematical theory of number representation and manipulation In Haggard P, Rossetti Y, Kawato M, editors, *Attention & performance XXII: Sensori-motor foundations of higher cognition*, pp. 527–574. Cambridge, MA: Harvard University Press.

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

Dehaene S, Cohen L (1997) Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33:219–250.

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

Dehaene S, Mehler J (1992) Cross-linguistic regularities in the frequency of number words. *Cognition* 43:1–29.

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

Della Puppa A, De Pallegrin S, d'Avella E, Gioffrè G, Munari M, Saladini M, Salillas E, Scienza R, Semenza C (2013) Right parietal cortex and calculation processing: Intraoperative functional mapping of multiplication and addition in patients affected by a brain tumor. *Journal of Neurosurgery* 119:1107–1111.

Diester I, Nieder A (2007) Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biology* 5:e294.

Diester I, Nieder A (2008) Complementary contributions of prefrontal neuron classes in abstract numerical categorization. *Journal of Neuroscience* 28:7737–7747.

Diester I, Nieder A (2010) Numerical values leave a semantic imprint on associated signs in monkeys. *Journal of Cognitive Neuroscience* 22:174–183.

Ditz HM, Nieder A (2015) Neurons selective to the number of visual items in the corvid songbird endbrain. *Proceedings of the National Academy of Sciences, USA* 112:7827–7832.

Ditz H, Fechner J, Nieder A (2022) Cell-type specific pallial circuits shape categorical tuning responses in the crow telencephalon. *Communications Biology* 5:269.

Duncan E, McFarland C (1980) Isolating the effects of symbolic distance and semantic congruity in comparative judgments: An additive-factors analysis. *Memory & Cognition* 8:612–622.

Eger E, Michel V, Thirion B, Amadon A, Dehaene S, Kleinschmidt A (2009) Deciphering cortical number coding from human brain activity patterns. *Current Biology* 19:1608–1615.

Eger E, Sterzer P, Russ M, Giraud A, Kleinschmidt A (2003) A supramodal number representation in human intraparietal cortex. *Neuron* 37:719–725.

Eichenbaum H, Yonelinas A, Ranganath C (2007) The medial temporal lobe and recognition memory. *Annual Review of Neuroscience* 30:123–152.

Eiselt AK, Nieder A (2013) Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. *Journal of Neuroscience* 33:7526–7534.

Ekstrom AD, Kahana MJ, Caplan JB, Fields TA, Isham EA, Newman EL, Fried I (2003) Cellular networks underlying human spatial navigation. *Nature* 425:184–188.

Fechner GT (1860) *Elemente der Psychophysik*, Vol. 2 Leipzig: Breitkopf und Härtel.

Fehr T, Code C, Herrmann M (2007) Common brain regions underlying different arithmetic operations as revealed by conjunct fMRI-BOLD activation. *Brain Research* 1172:93–102.

Feigenson L, Carey S (2003) Tracking individuals via object-files: Evidence from infants manual search. *Developmental Science* 6:568–584.

Feigenson L, Carey S, Hauser M (2002) The representations underlying infants' choice of more: Object-files versus analog magnitudes. *Psychological Science* 13:150–156.

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

Fias W, Lammertyn J, Reynvoet B, Dupont P, Orban G (2003) Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience* 15:47–56.

Fornaciai M, Park J (2021) Decoding of electroencephalogram signals shows no evidence of a neural signature for subitizing in sequential numerosity. *Journal of Cognitive Neuroscience* 33:1535—-1548.

Foster JJ, Vogel EK, Awh E (in press) Working memory as persistent neural activity In Kahana MJ, Wagner AD, editors, *Oxford Handbook of Human Memory*. Oxford: Oxford University Press.

Fried I, MacDonald K, Wilson C (1997) Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. *Neuron* 18:753–765.

Fried I, Rutishauser U, Cerf M, Kreiman G (2014) *Single neuron studies of the human brain: Probing cognition* MIT Press.

Funahashi S (2017) Working memory in the prefrontal cortex. *Brain Sciences* 7:49.

Fuster JaA G (1971) Neuron activity related to short-term memory. *Science* 173:652–654.

Fuster J (2015) *The prefrontal cortex* Academic Press, 5th edition.

Gallistel C (2017) Finding numbers in the brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373:20170119.

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

Gast H, Niediek J, Schindler K, Boström J, Coenen VA, Beck H, Elger CE, Mormann F (2016) Burst firing of single neurons in the human medial temporal lobe changes before epileptic seizures. *Clinical Neurophysiology* 127:3329—-3334.

Geary DC (2011) Cognitive predictors of achievement growth in mathematics: A 5-year longitudinal study. *Developmental Psychology* 47:1539–1552.

Geary DC, Bow-Thomas CC, Yao Y (1992) Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology* 54:372–391.

Geary DC, Hoard MK, Byrd-Craven J, DeSoto MC (2004) Strategy choice in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *Journal of Experimental Child Psychology* 88:121–151.

Gelman R, Gallistel CR (1978) *The child's understanding of number* Cambridge, MA: Harvard University Press.

Gilmore C, McCarthy S, Spelke E (2010) Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition* 115:394–406.

Goldman-Rakic P, Selemon L, Schwartz M (1984) Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience* 12:719–743.

Golman-Rakic PS (1995) Cellular basis of working memory. *Neuron* 14:477–485.

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

Grabner R, Ansari D, Koschutnig K, Reishofer G, Ebner F, C. N (2009) To retrieve or to calculate? left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia* 47:604—-608.

Green D, Swets J (1966) *Signal detection theory and psychophysics* New York: Wiley.

Gruber O, Indefrey P, Steinmetz H, Kleinschmidt A (2001) Dissociating neural correlates of cognitive components in mental calculation. *Cerebral Cortex* 11:350–359.

Halberda J, Feigenson L (2008) Developmental change in the acuity of the 'number sense': The approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology* 44:1457–1465.

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

Hartline H, Wagner H, Ratliff F (1956) Inhibition in the eye of *Limulus*. *Journal of General Physiology* 39:651—-673.

Harvey BM (2016) Quantity cognition: Numbers, numerosity, zero and mathematics. *Current Biology* 26:R419–R421.

Harvey B, Dumoulin S (2017) A network of topographic numerosity maps in human association cortex. *Nature Human Behavior* 1:36.

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

Hecht S (1999) Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition* 27:1097–1107.

Hermes D, Rangarajan V, Foster BL, Jean-Remi K, Kaşikçi I, Miller KJ, Parvizi J (2017) Electrophysiological responses in the ventral temporal cortex during reading of numerals and calculation. *Cerebral Cortex* 27:567–575.

Holloway I, Ansari D (2009) Mapping numerical magnitudes onto symbols. *Journal of Experimental Child Psychology* 103:17–29.

Hung Y, Hung D, Tzeng O, Wu D (2008) Flexible spatial mapping of different notations of numbers in chinese readers. *Cognition* 106:1441–1450.

Hyde D, Spelke E (2009) All numbers are not equal: An electrophysiological investigation of small and large number representations. *Journal of Cognitive Neuroscience* 21:1039–1053.

Ifrah G (1981) *Histoire universelle des chiffres* Paris: Seghers.

Isaacson J, Scanziani M (2011) How inhibition shapes cortical activity. *Neuron* 72:231–243.

Ison M, Mormann F, Cerf M, Koch C, Fried I, Quiroga R (2011) Selectivity of pyramidal cells and interneurons in the human medial temporal lobe. *Journal of Neurophysiology* 106:1713–1721.

Jacobs J, Weidemann CT, Miller JF, Solway A, Burke J, Wei XX, Suthana N, Sperling M, Sharan AD, Fried I, Kahana MJ (2013) Direct recordings of grid-like neuronal activity in human spatial navigation. *Nature Neuroscience* 16:1188–1190.

Jeneson A, Squire LR (2012) Working memory, long-term memory, and medial temporal lobe function. *Learning & Memory* 19:15–25.

Jevons WS (1871) The power of numerical discrimination. *Nature* 3:281–282.

Jordan NC, Hanich LB, Kaplan D (2003) Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology* 85:103–119.

Kaas JH (1997) Topographic maps are fundamental to sensory processing. *Brain Research Bulletin* 44:107–112.

Kamiński J, Rutishauser U (2020) Between persistently active and activity-silent frameworks: Novel vistas on the cellular basis of working memory. *Annals of the New York Academy of Sciences* 1464:64–75.

Kamiński J, Sullivan S, Chung JM, Ross IB, Mamelak AN, Rutishauser U (2017) Persistenly active neurons in human medial frontal and medial temporal lobe support working memory. *Nature Neuroscience* 20:590–601.

Kaufman E, Lord M, Reese T, Volkmann J (1949) The discrimination of visual number. *The American Journal of Psychology* 62:498–525.

King JR, Dehaene S (2014) Characterizing the dynamics of mental representations: The temporal generalization method. *Trends in Cognitive Sciences* 18:203–210.

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* 34:9857–9866.

Koch C, Ullman S (1985) Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology* 4:219–227.

Konkel A, Warren DE, Duff MC, Tranel D, Cohen NJ (2008) Hippocampal amnesia impairs all manner of relational memory. *Frontiers in Human Neuroscience* 2:15.

Kornblith S, Quian Quiroga R, Koch C, Fried I, Mormann F (2017) Persistent single-neuron activity during working memory in the human medial temporal lobe. *Current Biology* 27:1026–1032.

Krajcsi A, Szabó E, Mórocz IA (2013) Subitizing is sensitive to the arrangement of objects. *Experimental Psychology* 60:227–234.

Kreiman G, Koch C, Fried I (2000) Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience* 3:946–953.

Kucian K, Loenneker T, Dietrich T, Dosch M, Martin E, von Aster M (2006) Impaired neural networks for approximate calculation in dyscalculic children: a functional MRI study. *Behavioral and Brain Functions* 2:31.

LeFevre J, Bisanz J, Daley K, Buffone L, Greenham S, Sadesky G (1996) Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General* 125:284–306.

Leybaert J, van Cutsem M (2002) Counting in sign language. *Journal of Experimental Child Psychology* 81:482–501.

Libertus M, Woldorff M, Brannon E (2007) Electrophysiological evidence for notation independence in numerical processing. *Behavioral Brain Functions* 3:1–15.

Livingstone MS, Pettine WW, Srihasam K, Lee D (2014) Symbol addition by monkeys provides evidence for normalized quantity coding. *Proceedings of the National Academy of Sciences, USA* 111:6822–6827.

Logie RH, Baddeley AD (1987) Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:310–326.

Luck S, Vogel E (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390:279–281.

Lyons I, Ansari D, Bellock S (2015) Qualitatively different coding of symbolic and nonsymbolic numbers in the human brain. *Human Brain Mapping* 36:475–488.

Maloney E, Risko E, Preston F, Ansari D, Fugelsang J (2010) Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychologica* 134:154–161.

Mandler G, Shebo B (1982) Subitizing: An analysis of its component processes. *Journal of Experimental Psychology: General* 111:1–22.

Mansouri FA, Freedman DJ, Buckley MJ (2020) Emergence of abstract rules in the primate brain. *Nature Reviews Neuroscience* 21:595–610.

Marino A, Scholl B (2005) The role of closure in defining the 'objects' of object-based attention. *Perception & Psychophysics* 67:1140–1149.

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

Meck W, Church R (1983) A mode control model for counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes* 9:320–334.

Menon V (2015) Arithmetic in the child and adult brain In Cohen Kadosh R, Dowker A, editors, *The Oxford Handbook of Mathematical Cognition*. Oxford: Oxford University Press.

Menon V (2016) Memory and cognitive control circuits in mathematical cognition and learning. *Progress in Brain Research* 227:159–186.

Menon V, Rivera S, White C, Glover G, Reiss A (2000) Dissociating prefrontal and parietal cortex activation during arithmetic processing. *NeuroImage* 12:357–365.

Merten K, Nieder A (2009) Compressed scaling of abstract numerosity representations in adult humans and monkeys. *Journal of Cognitive Neuroscience* 21:333–346.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience* 24:167–202.

Miller EK, M. L, Bastos AM (2018) Working memory 2.0. *Neuron* 100:463–475.

Mongillo G, Barak O, Tsodyks M (2008) Synaptic theory of working memory. *Science* 319:1543–1546.

Mormann F, Dubois J, Kornblith S, Milosavljevic M, Cerf M, Ison M, Tsuchiya N, Kraskov A, Quian Quiroga R, Adolphs R, Fried I, Koch C (2011) A category-specific response to animals in the right human amygdala. *Nature Neuroscience* 14:1247–1249.

Mosher C, Wei Y, Kamiński J, Nandi A, Mamelak A, Anastassiou C, Rutishauser U (2020) Cellular classes in the human brain revealed in vivo by heartbeat-related modulation of the extracellular action potential waveform. *Cell Reports* 30:3536–3551.e6.

Moyer RS, Landauer TK (1967) Time required for judgements of numerical inequality. *Nature* 215:1519–1520.

Mukamel R, Ekstrom AD, Kaplan J, Iacoboni M, Fried I (2010) Single-neuron responses in humans during execution and observation of actions. *Current Biology* 20:750–756.

Murray JD, Bernacchia A, Roy NA, Constantinidis C, Romo R, Wang XJ (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences, USA* 114:394–399.

Naccache L, Dehaene S (2001) The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal cortex. *Cerebral Cortex* 11:966–974.

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

Nieder A (2009) Prefrontal cortex and the evolution of symbolic reference. *Current Opinion in Neurobiology* 19:99–108.

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

Nieder A (2016a) The neuronal code for number. *Nature Reviews Neuroscience* 17:366–382.

Nieder A (2016b) Representing something out of nothing: The dawning of zero. *Trends in Cognitive Sciences* 20:830–842.

Nieder A (2021) Neuroethology of number sense across the animal kingdom. *Journal of Experimental Biology* 224.6:jeb218289.

Nieder A, Dehaene S (2009) Representation of number in the brain. *Annual Review of Neuroscience* 32:185–208.

Nieder A, Diester I, Tudusciuc O (2006) Temporal and spatial enumeration processes in the primate parietal cortex. *Science* 313:1431–1435.

Nieder A, Freedman DJ, Miller EK (2002) Representation of the quantity of visual items in the primate prefrontal cortex. *Science* 297:1708–1711.

Nieder A, Merten K (2007) A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *The Journal of Neuroscience* 27:5986–5993.

Nieder A, Miller EK (2003) Coding of cognitive magnitude: Compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* 37:149–157.

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

Nuerk HC, Wood G, Willmes K (2005) The universal SNARC effect: The association between number magnitude and space is amodal. *Experimental Psychology* 52:187–194.

Olson IR, Page K, Sledge Moore K, Chatterjee A, Verfaellie M (2006) Working memory for conjunctions relies on the medial temporal lobe. *The Journal of Neuroscience* 26:4596–4601.

Opris I, Santos L, Gerhardt G, Song D, Berger T, Hampson R, , Deadwyler S (2015) Distributed encoding of spatial and ob- ject categories in primate hippocampal microcircuits. *Frontiers in Neuroscience* 9:317.

Owen AM, McMillan KM, Laird AR, Bullmore E (2005) N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping* 25:46–59.

Park J, De Wind N, Woldorff M, Brannon E (2016) Rapid and direct encoding of numerosity in the visual stream. *Cerebral Cortex* 26:748–763.

Peters L, De Smedt B (2018) Arithmetic in the developing brain: A review of brain imaging studies. *Developmental Cognitive Neuroscience* 30:265–279.

Piazza M, Fumarola A, Chinello A, Melcher D (2011) Subitizing reflects visuo-spatial object individuation capacity. *Cognition* 121:147—-153.

Piazza M, Izard V, Pinel P, Le Bihan D, Dehaene S (2004) Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44:547–555.

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

Piazza M (2010) Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences* 14:542–551.

Piazza M, Izard V (2009) How humans count: Numerosity and the parietal cortex. *Neuroscientist* 15:261–273.

Piazza M, Mechelli A, Butterworth B, Price C (2002) Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage* 15:435—-446.

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

Pinel P, Piazza M, Le Bihan D, Dehaene S (2004) Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. *Neuron* 41:983–993.

Pinel P, Dehaene S, Rivière D, Le Bihan D (2001) Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage* 14:1013–1026.

Pinheiro-Chagas P, Daitch A, Parvizi J, Dehaene S (2018) Brain mechanisms of arithmetic: A crucial role for ventral temporal cortex. *Journal of Cognitive Neuroscience* 30:1757–1772.

Plaisier MA, Bergmann Tiest WM, Kapper AM (2009) One, two, three, many – Subitizing in active touch. *Acta Psychologica* 131:163–170.

Pomè A, Thompson D, Burr DC, Halberda J (2020) Location- and object-based attention enhance number estimation. *Attention, Perception & Psychophysics* 83:7–17.

Qi XL, Meyer T, Stanford TR, Constantinidis C (2011) Changes in prefrontal neuronal activity after learning to perform a spatial working memory task. *Cerebral Cortex* 21:2722–2732.

Qin S, Cho S, Chen T, Rosenberg-Lee M, Geary D, Menon V (2014) Hippocampal-neocortical functional reorganization underlies children's cognitive development. *Nature Neuroscience* 17:1263–1269.

Quian Quiroga R, Kraskov A, Koch C, Fried I (2009) Explicit encoding of multimodal percepts by single neurons in the human brain. *Current Biology* 19:1308–1313.

Quian Quiroga R (2012) Concept cells: The building blocks of declarative memory functions. *Nature Review Neuroscience* 13:587–597.

Quian Quiroga R, Reddy L, Kreiman G, Koch C, Fried I (2005) Invariant visual representation by single neurons in the human brain. *Nature* 435:1102–1107.

Railo H, Koivisto, M. R, A., Hannula MM (2008) The role of attention in subitizing. *Cognition* 107:82—-104.

Ramirez-Cardenas A, Moskaleva M, Nieder A (2016) Neuronal representation of numerosity zero in the primate parieto-frontal number network. *Current Biology* 26:1285–1294.

Repp BH (2007) Perceiving the numerosity of rapidly occurring auditory events in metrical and nonmetrical contexts. *Perception & Psychophysics* 69:529–543.

Revkin S, Piaza M, Izard V, Cohen L, Dehaene S (2008) Does subitizing reflect numerical estiamtion? *Psychological Science* 19:607–614.

Riggs KJ, Ferrand L, Lancelin D, Fryziel L, Dumur G, Simpson A (2006) Subitizing in tactile perception. *Psychological Science* 17:271–272.

Roggeman C, Fias W, Verguts T (2010) Salience maps in parietal cortex: Imaging and computational modeling. *NeuroImage* 52:1005–1014.

Roitman J, Brannon E, Platt M (2007) Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS Biology* 5:e208.

Romo R, Salinas E (2003) Flutter discrimination: Neural codes, perception, memory and decision making. *Nature Reviews Neuroscience* 4:203–218.

Rose NS (2020) The dynamic-processing model of working memory. *Current Directions in Psychological Science* 29:378–387.

Rose N, LaRoque J, Rigall A, Gosseries O, Starrett M, Meyering E, Postle B (2016) Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 22:265–266.

Rosenberg-Lee M, Chang TT, Young CB, Wu S, Menon V (2011) Functional dissociations between four basic arithmetic operations in the human posterior parietal cortex: A cytoarchitectonic mapping study. *Neuropsychologia* 49:2592–2608.

Roux FE, Boukhatem L, Draper L, Sacko O, Démonet JF (2009) Cortical calculation localization using electrostimulation. *Journal of Neurosurgery* 110:1291–1299.

Rumbaugh D, Savage-Rumbaugh S, Hegel M (1987) Summation in the chimpanzee (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes* 13:107–115.

Rutishauser U, Reddy L, Mormann F, Sarnthein J (2021) The architecture of human memory: Insights from human single-neuron recordings. *The Journal of Neuroscience* 41:883–890.

Rykhlevskaia E, Uddin L, Kondos L, Menon V (2009) Neuroanatomical correlates of developmental dyscalculia: Combined evidence from morphometry and tractography. *Frontiers in Human Neuroscience* 3:51.

Sawamura H, Shima K, Tanji J (2002) Numerical representation for action in the parietal cortex of the monkey. *Nature* 415:918–922.

Sawamura H, Shima K, Tanji J (2010) Deficits in action selection based on numerical information after inactivation of the posterior parietal cortex. *Journal of Neurophysiology* 104:902–910.

Saxena S, Cunningham JP (2019) Towards the neural population doctrine. *Current Opinion in Neurobiology* 33:103–111.

Sekuler R, Mierkiewicz D (1977) Children's judgment of numerical inequality. *Child Development* 48:630–633.

Semenza C, Salillas E, De Pallegrin S, Della Puppa A (2017) Balancing the 2 hemispheres in simple calculation: Evidence from direct cortical electrosimulation. *Cerebral Cortex* 27:4806–4814.

Sengupta R, Surampudi BR, Melcher D (2014) A visual sense of number emerges from the dynamics of a recurrent on-center ooff-surround neural network. *Brain Research* 1582:114–124.

Shum J, Hermes D, Foster BL, Dastjerdi M, Rangarajan V, Winawer J, Miller KJ, Parvizi J (2013) A brain area for visual numerals. *The Journal of Cognitive Neuroscience* 33:6709–6715.

Siegler RS (1987) The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General* 116:250–264.

Siegler R (1996) *Emerging minds: The process of change in children's thinking* Oxford: Oxford University Press.

Spaak E, Watanabe K, Funahashi S, Stokes MG (2017) Stable and dynamic coding for working memory in primate prefrontal cortex. *Journal of Neuroscience* 37:6503–6516.

Spelke ES, Kinzler KD (2007) Core knowledge. *Developmental Science* 10:89–96.

Squire LR (2017) Memory for relations in the short term and the long term after medial temporal lobe damage. *Hippocampus* 27:608–612.

Squire LR, Zola-Morgan S (1991) The medial temporal lobe memory system. *Science* 253:1380–1386.

Squire L, Stark C, Clark R (2004) The medial temporal lobe. *Annual Review of Neuroscience* 27:279–306.

Sreenivasan KK, Curtis CE, D'Esposito M (2014) Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences* 18:82–89.

Stanislaw H, Todorov N (1999) Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers* 31:137–149.

Starkey P, Spelke E, Gelman R (1990) Numerical abstraction by human infants. *Cognition* 36:97–127.

Starr A, Libertus M, Brannon E (2013) Number sense in infancy predicts mathematical abilities in childhood. *Proceedings of the National Academy of Sciences, USA* 110:18116–18120.

Stoet G, Snyder LH (2004) Single neurons in posterior parietal cortex of monkeys encode cognitive set. *Neuron* 42:1003–1012.

Stoianov I, Zorzi M (2012) Emergence of a 'visual number sense' in hierarchical generative models. *Nature Neuroscience* 15:194–196.

Stokes MG, Kusunoki M, Sigala N, Nili H, Gaffan D, Duncan J (2013) Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78:364–375.

Supekar K, Swigart A, Tenison C, Jolles D, Rosenberg-Lee M, Fuchs L, Menon V (2013) Neural predictors of individual differences in response to math tutoring in primary-grade school children. *Proceedings of the National Academy of Sciences, USA* 110:8230–8235.

Suzuki W (2009) Comparative analysis of the cortical afferents, intrinsic projections, and interconnections of the parahippocampal region in monkeys and rats In Gazzaniga M, editor, *The Cognitive Neurosciences IV*, pp. 659–674. Cambridge: MIT Press.

Szkudlarek E, Brannon EM (2017) Does the approximate number system serve as a foundation for symbolic mathematics. *Language, Learning and Development* 13:171–190.

ten Hoopen G, Vos J (1979) Effect on numerosity judgment of grouping of tones by auditory channels. *Perception & Psychophysics* 26:374–380.

Thivierge JP, Marcus GF (2007) The topographic brain: From neural connectivity to cognition. *Trends in Neuroscience* 30:251–259.

Trick LM, Pylyshyn ZW (1994) Why are small and large numbers enumerated differently? a limited-capacity preattentive stage in vision. *Psychological Review* 101:80–102.

Tschentscher N, Hauk O (2014) How are things adding up? neural differences between arithmetic operations are due to general problem solving strategies. *Neuroimage* 92:369—-380.

Tsouli A, Harvey BM, Hofstetter S, Cai Y, van der Smagt MJ, te Pas SF, Dumoulin SO (2022) The role of neural tuning in quantity perception. *Trends in Cognitive Sciences* 26:11–24.

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, USA* 104:14513–14518.

Tudusciuc O, Nieder A (2009) Contributions of primate prefrontal and posterior parietal cortices to length and numerosity representation. *Journal of Neurophysiology* 101:2984–2994.

Tulving E, Markowitsch HJ (1998) Episodic and declarative memory: Role of the hippocampus. *Hippocampus* 8:198–204.

Vallentin D, Bongard S, Nieder A (2012) Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. *Journal of Neuroscience* 32:6621–6630.

van Marle K, Scholl BJ (2003) Attentive tracking of object versus substances. *Psychological Sciences* 14:498–504.

van Oeffelen MP, Vos PG (1982) A probabilistic model for the discrimination of visual number. *Perception & Psychophysics* 32:163–170.

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

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, USA* 110:11187–11192.

Viswanathan P, Nieder A (2015) Differential impact of behavioral relevance on quantitative coding in primate frontal and parietal neurons. *Current Biology* 25:1259–1269.

Wallis JD, Anderson KC, Miller EK (2001) Single neurons in prefrontal cortex encode abstract rules. *Nature* 411:953–956.

Wallis JD, Miller EK (2003) From rule to response: Neuronal processes in the premotor and prefrontal cortex. *Journal of Neurophysiology* 90:1790–1806.

Wark B, Lundstrom BN, Fairhall A (2007) Sensory adaptation. *Current Opinion in Neurobiology* 17:423–429.

Washburn D, Rumbaugh D (1991) Ordinal judgments of numerical symbols by macaques (*Macaca mulatta*). *Psychological Science* 2:190–193.

Weber EH (1850) Der Tastsinn und das Gemeingefühl In Wagner R, editor, *Handwörterbuch der Physiologie*, Vol. 3, pp. 481–588. Braunschweig: Vieweg.

Whalen J, Gallistel C, Gelman R (1999) Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science* 10:130–137.

White IM, Wise SP (1999) Rule-dependent neuronal activity in the prefrontal cortex. *Experimental Brain Research* 126:315–335.

Wilkey ED, Ansari D (2019) Challenging the neurobiological link between number sense and symbolic numerical abilities. *Annals of the New York Academy of Sciences* 1464:76–98.

Wolff M, Ding J, Myers N, Stokes MG (2015) Revealing hidden states in visual memory using electroencephalography. *Frontiers in Systems Neuroscience* 9:1–12.

Wolff M, Jochim J, Akyurek E, Stokes MG (2017) Dynamic hidden states underlying working-memory-guided behaviour. *Nature Neuroscience* 20:864–871.

Woodruff G, Premack D (1981) Primative mathematical concepts in the chimpanzee: Proportionality and numerosity. *Nature* 293:568–570.

Wynn K (1990) Children's understanding of counting. *Cognition* 36:155–193.

Wynn K (1992) Addition and subtraction by human infants. *Nature* 358:749–750.

Xu F (2003) Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition* 89:B15–B25.

Xu F, Spelke E (2000) Large number discrimination in 6-month-old infants. *Cognition* 74:B1–B11.

Xu F, Arriaga RI (2007) Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology* 25:103–108.

Yuste R (2015) From the neuron doctrine to neural networks. *Nature Reviews Neuroscience* 16:487–497.

Zhou X, Shen C, Li L, Li D, Cui J (2016) Mental numerosity line in the human's approximate number system. *Experimental Psychology* 63:169–179.

Zucker RS, Regehr WG (2002) Short-term synaptic plasticity. *Annual Review of Neurophysiology* 64:355–405.

Part II.
Individual Publications

Statement of Contributions

This thesis comprises three publications which are summarized and discussed in Part I of this dissertation.

1. **Kutter E.F.**, Bostroem J., Elger C.E., Mormann F., & Nieder A. (2018) Single Neurons in the Human Brain Encode Numbers. *Neuron* **100(3)**: 1–9. DOI: 10.1016/j.neuron.2018.08.036

   A. Nieder and F. Mormann designed the study. C.E. Elger and F. Mormann recruited patients; J. Bostroem and F. Mormann implanted the electrodes. I and F. Mormann collected the data. I analysed and interpreted the data with contributions from A. Nieder and F. Mormann. I wrote the manuscript together with A. Nieder and F. Mormann.

2. **Kutter E.F.**, Bostroem 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. DOI: 10.1016/j.cub.2022.01.054

   A. Nieder and F. Mormann designed the study. C.E. Elger and F. Mormann recruited patients; J. Bostroem and F. Mormann implanted the electrodes. I and F. Mormann collected the data. I analysed and interpreted the data with contributions from A. Nieder and F. Mormann. I wrote the manuscript together with F. Mormann and A. Nieder.

3. **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**: 1998–2007. DOI: 10.1038/s41562-023-01709-3

   A. Nieder and F. Mormann designed the study. R. Surges and F. Mormann recruited patients; V. Borger and F. Mormann implanted the electrodes. I and G. Dehnen collected the data. I and A. Nieder analysed and interpreted the data with contributions from F. Mormann. I wrote the manuscript together with A. Nieder and F. Mormann.

Publication 1: Single Neurons in the Human Brain Encode Numbers

Kutter E.F., Bostroem J., Elger C.E., Mormann F., & Nieder A. (2018) Single Neurons in the Human Brain Encode Numbers. *Neuron* **100(3)**: 1–9. DOI: 10.1016/j.neuron.2018.08.036

Article

Neuron

Single Neurons in the Human Brain Encode Numbers

Highlights

- Single neurons in the human medial temporal lobe (MTL) encode numerical information
- Numerosity and abstract numerals are encoded by distinct neuronal populations
- Numerosity representation shows a distance effect; numerals are encoded categorically
- Representation of symbolic numerals may evolve from numerosity representations

Authors

Esther F. Kutter, Jan Bostroem, Christian E. Elger, Florian Mormann, Andreas Nieder

Correspondence

florian.mormann@ukbonn.de (F.M.), andreas.nieder@uni-tuebingen.de (A.N.)

In Brief

Kutter et al. show how nonsymbolic and symbolic numerical quantity is encoded in the human brain by neurons of the medial temporal lobe. The data support the hypothesis that high-level human numerical abilities are rooted in biologically determined mechanisms.

Kutter et al., 2018, Neuron *100*, 1–9
November 7, 2018
https://doi.org/10.1016/j.neuron.2018.08.036

CellPress

Please cite this article in press as: Kutter et al., Single Neurons in the Human Brain Encode Numbers, Neuron (2018), https://doi.org/10.1016/j.neuron.2018.08.036

Neuron
Article

Single Neurons in the Human Brain Encode Numbers

Esther F. Kutter,[1,2] Jan Bostroem,[3] Christian E. Elger,[1] Florian Mormann,[1,4,*] and Andreas Nieder[2,4,5,*]
[1]Department of Epileptology, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany
[2]Animal Physiology Unit, Institute of Neurobiology, University of Tübingen, 72076 Tübingen, Germany
[3]Department of Neurosurgery, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany
[4]Senior author
[5]Lead Contact
*Correspondence: florian.mormann@ukbonn.de (F.M.), andreas.nieder@uni-tuebingen.de (A.N.)
https://doi.org/10.1016/j.neuron.2018.08.036

SUMMARY

Our human-specific symbolic number skills that underpin science and technology spring from nonsymbolic set size representations. Despite the significance of numerical competence, its single-neuron mechanisms in the human brain are unknown. We therefore recorded from single neurons in the medial temporal lobe of neurosurgical patients that performed a calculation task. We found that distinct groups of neurons represented either nonsymbolic or symbolic number, but not both number formats simultaneously. Numerical information could be decoded robustly from the population of neurons tuned to nonsymbolic number and with lower accuracy also from the population of neurons selective to number symbols. The tuning characteristics of selective neurons may explain why set size is represented only approximately in behavior, whereas number symbols allow exact assessments of numerical values. Our results suggest number neurons as neuronal basis of human number representations that ultimately give rise to number theory and mathematics.

INTRODUCTION

Numbers are fundamental to science and technology. Despite counting and arithmetic requiring years of training, the origins of our symbolic number capabilities are deeply rooted in our ancestry (Dehaene, 1997). Human adults without formal education (Gordon, 2004; Pica et al., 2004), pre-linguistic human infants (Wynn, 1992; Xu and Spelke, 2000), and nonhuman animals (Brannon and Terrace, 1998; Scarf et al., 2011) can approximately estimate numerosity, the number of items in a set. These intuitive nonsymbolic capabilities are harnessed and qualitatively transformed by children when they begin to learn symbolic counting and mathematics in school (Halberda et al., 2008; Gilmore et al., 2010; Starr et al., 2013). This intimate relationship between set size estimation and precise counting suggests that symbolic arithmetic abilities build on nonsymbolic numerical capacities.

Studies in humans (Piazza et al., 2007; Arsalidou and Taylor, 2011) and nonhuman primates (Nieder, 2016) indicated parts of the parietal and prefrontal cortices as a core number system that processes nonsymbolic and symbolic numerical magnitude. However, the wider cortical number network also incorporates areas of the medial temporal lobe (MTL) (Menon, 2016), such as the hippocampus, parahippocampal cortex, entorhinal cortex, and amygdala. The MTL comprises highly associative brain areas that are directly and reciprocally connected with the frontal number network (Goldman-Rakic et al., 1984), and human MTL neurons are known for their selectivity to abstract categories (Quiroga et al., 2005; Mormann et al., 2011). Functional imaging studies in humans showed that the hippocampal system—among many other functions outside of the number domain—is also involved in learning to count and arithmetic skill acquisition, specifically during childhood (De Smedt et al., 2011; Supekar et al., 2013; Qin et al., 2014). Hippocampal-frontal circuit reorganization plays an important role in children's shift from effortful counting to efficient memory-based solving of mathematical problems (Menon, 2016).

As a neuronal correlate of numerosity representations, electrophysiological recordings from the association cortex of monkeys showed neurons that are tuned to a specific preferred numerosity of visual and auditory items. Such number neurons have also been postulated by neural network models (Dehaene and Changeux, 1993; Verguts and Fias, 2004). In humans, number neurons have been suggested based on blood flow changes in functional imaging studies (Piazza et al., 2004; Jacob and Nieder, 2009a), as well as the combined synaptic mass signals from hundreds of neurons measured with electrocorticography (ECoG) (Daitch et al., 2016). Despite the progress that has been made using functional imaging and ECoG recordings, the mechanism of how single neurons, the anatomical and functional units of the brain, encode nonsymbolic or symbolic numerical information in humans remains unknown. We addressed this question and recorded from single neurons in the MTL of neurosurgical patients that performed a calculation task and were implanted with intracranial electrodes (Fried et al., 1997; Kreiman et al., 2000; Reber et al., 2017).

RESULTS

Participants performed simple sequential addition and subtraction tasks using a computer display (Figure 1A). Task involvement ensured that numbers shown as operands were

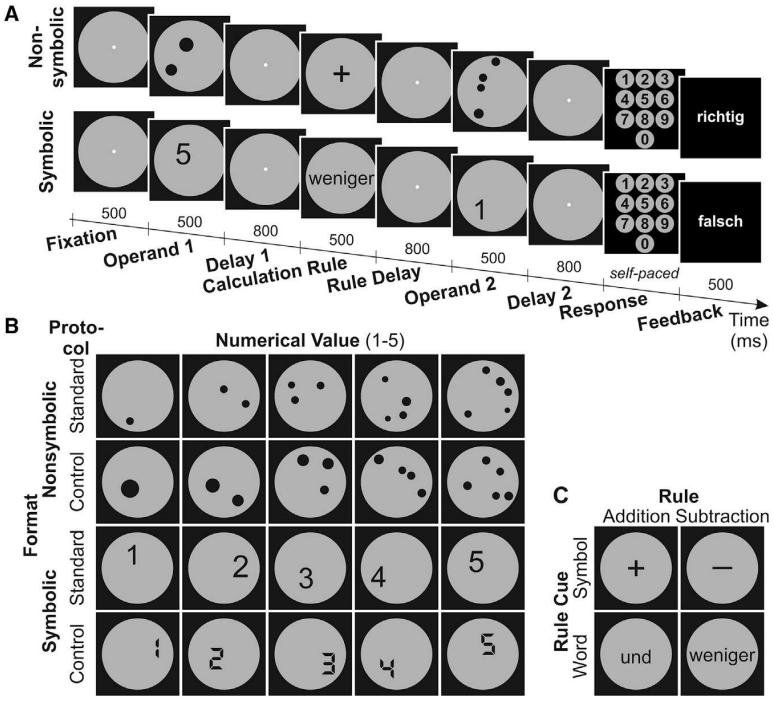


Figure 1. Behavioral Task and Example Stimuli

(A) Experimental design of the calculation task. After visual fixation on the screen, the first number (operand 1) was followed by a brief delay, after which the addition or subtraction rule was presented, followed in turn by a delay and then the second number (operand 2). After another brief delay, subjects were required to indicate the calculated result (ranging from 0 to 9) on a number panel.

(B) Example operand 1 stimuli for the nonsymbolic and symbolic format for standard and control protocols.

(C) Example stimuli for the different calculation rules indicated by arithmetic symbols ("+" and "−") and written words ("und" [add] and "weniger" [subtract]), respectively.

consciously processed. Numerical values of the operands ranged from 1 to 5. In half of the shuffled trials, the numerical values were presented nonsymbolically as the number of randomly placed dots in an array (numerosity). In the other half, Arabic numerals were shown as symbolic number representations. Both nonsymbolic and symbolic numbers were shown in standard and control displays in order to control for low-level visual features (Figure 1B; see Supplemental Information). Arithmetic symbols or words were applied for addition and subtraction instructions (Figure 1C). Average performance of all participants was close to ceiling for all tested quantities and calculations (performance range 90.3%–99.8%).

We recorded from 585 single neurons in the medial temporal lobes (153 amygdala, 126 parahippocampal cortex, 107 entorhinal cortex, and 199 hippocampus) of nine human subjects performing the calculation tasks. In order to explore pure number representations, and to avoid confounds with cognitive factors later in the task, we focus on the presentation of the first operand (operand 1) and the subsequent working memory phase (delay 1); the remaining task phases are considered toward the end of the results. Random presentation of either the nonsymbolic or symbolic format from trial to trial allowed us to investigate an individual neuron's responses to each of the formats individually, but also to both formats, in an unbiased way.

Single Neurons Encode Nonsymbolic Number

When the participants calculated with numerosities (nonsymbolic format), a substantial proportion of the tested neurons (16%; $p << 0.001$ in binomial test; $p_{chance} = 0.01$; see also Supplemental Information for verification with shuffled data) showed activity that varied exclusively with the number of items during operand 1 presentation and the working memory delay 1 that followed, irrespective of the dot array layout (2-factor sliding-window ANOVA, with factors "numerical value" × "protocol"; $\alpha = 0.01$; Figure S1, left). Four of such numerosity-selective neurons are shown in Figure 2, left column. Each cell is tuned to numerosity; it shows peak activity for one of the numerosities, its preferred numerosity, and a systematic decrease of activity the more the number of items deviates from the preferred value. The highest fraction of such numerosity-selective neurons in the MTL was found in the parahippocampal cortex (29%), followed by the hippocampus (18%; Figure 3, upper columns). The selective neurons' preference covered the entire tested range of numerosities, albeit with most neurons preferring numerosity "five" (Figure 4A, left). The proportion of neurons selective to nonsymbolic number for each subject is shown in Table S1. Firing rates were generally low in the MTL, but the firing rates of numerosity-selective neurons were significantly higher compared to the non-selective neurons ($p < 0.0001$; Mann-Whitney U test; Figure S2, left).

Average tuning curves were calculated by averaging the normalized activity for all numerosity-selective neurons that preferred a given numerosity. Neural activity formed overlapping tuning functions with progressively reduced activity as distance from the preferred quantity increased (Figure 4B, left). To compare the decay of activity from the preferred quantity across all neurons tuned to preferred numerosities 1–5, we plotted the normalized firing rates as a function of absolute numerical distance from the preferred numerosity. For example, the normalized firing rate to numerosity 2 and 4 of a cell tuned to numerosity 3 (3 therefore is absolute numerical distance 0) were plotted at absolute numerical distance 1. The pooled function for all selective neurons compared to a function from random tuning curves is shown in Figure 4C, left. On average, activity dropped off progressively with numerical distance across all preferred numerosities, an effect that is not observed for random tuning curves. This

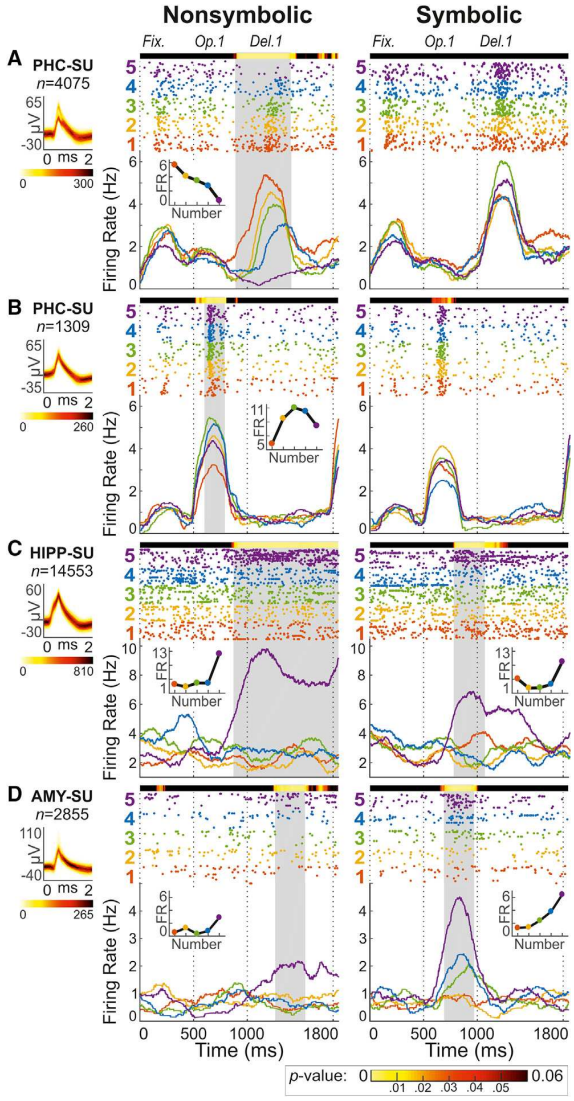


Figure 2. Neural Responses of Number-Selective Neurons during Presentation of Operand 1 and Delay 1

Responses of four example neurons to both nonsymbolic numerosities (left column) and symbolic numerals (right column). The left panels depict a density plot of the recorded action potentials (color darkness indicates number of overlapping wave forms according to color scale at the bottom). Panels show single-cell response rasters for many repetitions of the format (each dot represents an action potential) and averaged instantaneous firing rates below. The first 500 ms represent the fixation period. Colors correspond to the five different operand 1 values. Gray shaded areas represent significant number discrimination periods according to the sliding-window ANOVA (color-coded p values above each panel). Insets show the number tuning functions.

(A and B) Two parahippocampal neurons only responsive to nonsymbolic number with preferred numerosity 1 (A) and 3 (B).

(C and D) Hippocampal neuron #1 (C) and neuron #2 (D) responding to both nonsymbolic and symbolic number 5.

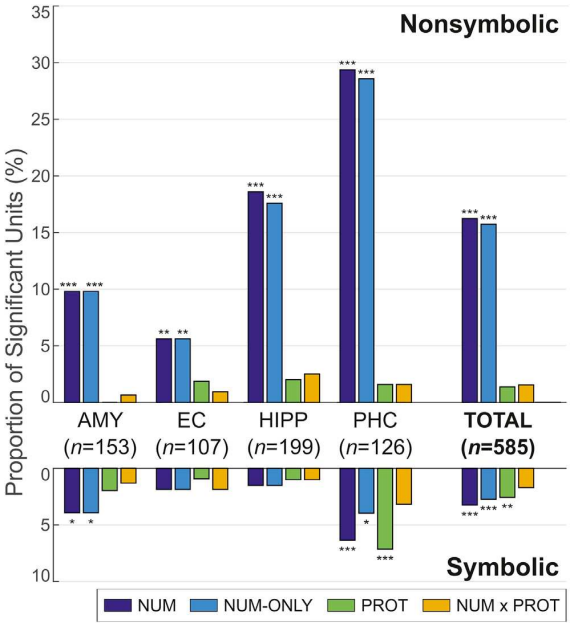


Figure 3. Neuronal Selectivity of MTL Single Units

Proportions of single units with significant main effects for "numerical value" (NUM: 1–5) or "protocol" (PROT: standard and control) and interactions (NUM × PROT) in a 2-factor ANOVA evaluated at $\alpha = 0.01$, separately for each format and MTL region (AMY, amygdala; EC, entorhinal cortex; HIPP, hippocampus; PHC, parahippocampal cortex). All analyses refer to exclusively number-selective (NUM-ONLY) units, i.e., neurons with an effect for numerical value but no concurrent effects for protocol or interaction. Numbers of significant neurons were subjected to a Bonferroni-corrected (n = 4) binomial test; asterisks indicate significance (*p < 0.05, **p < 0.01, and ***p < 0.001).

finding reflects a neuronal correlate of the well-known "numerical distance effect," the behavioral observation that discrimination progressively enhances as numerical distance between two quantities increases (Buckley and Gillman, 1974; Merten and Nieder, 2009). A cross-validation analysis (see Supplemental Information) yielded high reproducibility of preferred numerosity for the population of numerosity-selective units (average correlation coefficient $r = 0.83$; $p < 0.0001$), indicating that the preferred numerosity of the neurons was reliable and robust.

Single-Cell Responses to Symbolic Number

When participants calculated with Arabic numerals (symbolic format), a smaller but significant proportion of the recorded neurons (3%; $p < 0.001$ in binomial test; $p_{chance} = 0.01$) responded selectively to numerals during operand 1 presentation and the subsequent working memory delay 1 (2-factor sliding-window ANOVA, with factors numerical value × protocol; $\alpha = 0.01$; Figure S1, right). The highest fraction of such numeral-selective neurons in the MTL was again found in the parahippocampal cortex (6%), followed by the amygdala (4%; Figure 3, lower columns). Six numeral-selective neurons (1% of all neurons) were also tuned to nonsymbolic number, which was more than

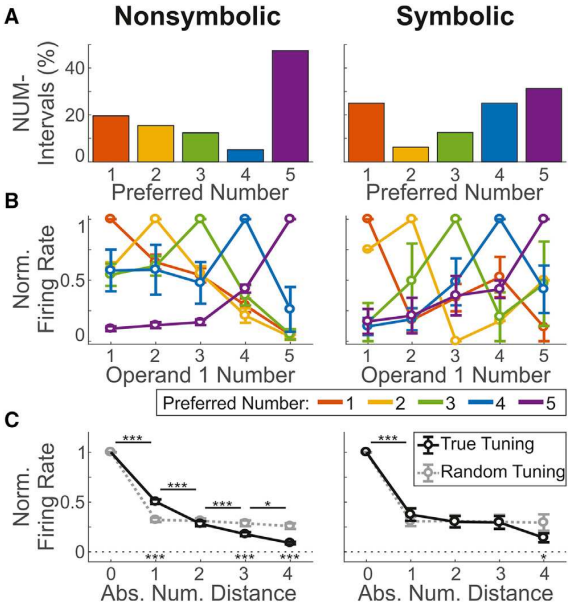


Figure 4. Tuning Properties of Number-Selective Neurons

(A) Frequency distribution of the preferred number of neurons tuned to numerosity (left) and numerals (right).

(B) Average tuning curves of neurons tuned to the five numerosities (left) and numerals (right).

(C) Averaged normalized activity across all preferred numerosities (left) and numerals (right) as a function of absolute numerical distance (black line). Asterisks above the graph represent significant differences between responses to adjacent numerical distances; asterisks below the dashed line indicate significant differences between recorded and random tuning curves (*p < 0.05 and ***p < 0.001). Error bars denote SEM.

expected by chance ($p < 0.05$ in binomial test; $p_{chance} = 0.16 \times 0.03 = 0.005$, or 0.5%). Of these, four neurons had identical preferred numerical values for nonsymbolic and symbolic number. This correlation did not reach significance (Figure S3A), possibly due to the small sample size. Next, we investigated whether the preferred numbers of units tuned to nonsymbolic numerosity might be correlated with their (non-significant) tuning to symbolic numerals (Figure S3B) and vice versa (Figure S3C). Neither correlation reached significance, indicating that numerosity and abstract numerals are encoded by two largely distinct neuronal populations. Two neurons tuned to the same value in both nonsymbolic and symbolic formats are depicted in Figures 2C and 2D. The neuron in Figure 2C as well as the neuron in Figure 2D showed maximum responses to quantity 5 in both the nonsymbolic and symbolic format. In contrast, the two neurons shown in Figures 2A and 2B were only significantly tuned to dot numerosities, but not to numerals. Again, a cross-validation analysis confirmed the reliability of the preferred numeral determination (average correlation coefficient $r = 0.57$; $p < 0.05$). The proportion of neurons selective to symbolic number for each subject is shown in Table S1. As for nonsymbolic number, the firing rates of numeral-selective neurons were significantly higher compared to the non-selective neurons ($p < 0.01$; Mann-Whitney U test; Figure S2, right).

Overall, the numeral-selective neurons' preference covered the entire range of numbers 1–5 (Figure 4A, right), and their normalized activity for each preferred numeral formed overlapping tuning functions (Figure 4B, right). The decline of activity from the preferred to the nonpreferred numerals was brisk and categorical, with only a mild progressive decrease with numerical distance, hardly showing a neuronal numerical distance effect (Figure 4C, right). At absolute numerical distance 1, the normalized firing rates obtained for symbolic number (n = 16) were significantly lower compared to nonsymbolic number (n = 92; $p < 0.05$; t test), indicating higher selectivity for (or sharper tuning to) symbolic number. When comparing the neuronal latencies to reach number-selectivity, neurons tuned to nonsymbolic (990 ms) and symbolic number (880 ms) did not differ significantly ($p = 0.23$; Mann-Whitney U test).

Neuronal Population Coding

So far, our data suggest two main findings at the level of individual neurons. First, the representation of nonsymbolic number was abundant and comparable to the core number network in nonhuman primates (Nieder et al., 2002, 2006; Nieder and Miller, 2004), whereas the representation of symbolic numbers was sparse in the MTL. Second, neurons responsive to nonsymbolic or symbolic number formats are largely segregated in the MTL; abstract neurons that encode the same numerical value in both nonsymbolic and symbolic formats were rarely found.

We therefore explored how the two populations of numerosity-selective and numeral-selective neurons encode numerical values. To evaluate the neuronal populations' information carried about number, we first trained a multi-class support vector machine (SVM) classifier to discriminate numerical values based on the spiking activity of selective MTL neurons (see Supplemental Information). After training, the classifier was tested with novel data from the same neuronal population to explore how well it could predict number categories based on the information extracted from trials used for classifier training. Initially, we performed a temporal cross-training classification to assess the classifier's accuracy in identifying the correct numerical values when tested on the activity from a given time period after being trained on other time periods of the trials. With a chance performance of 20% (for five classes), the classifier accuracy was significantly higher for both nonsymbolic and symbolic number throughout the operand 1 and delay 1 phases, albeit with better performance during the nonsymbolic-format trials (Figures 5A and 5B).

Next, we trained and tested the classifier on the firing rates of each neuron obtained by averaging across the time window that had turned out significant in the cross-training classification. The resulting confusion matrices show robust accuracy (65.6% ± 2.5%) for the five numerosities in the nonsymbolic format represented by the diagonal (Figure 5C, left). The probability of misclassification of trials increased the closer two classes were in the numerical space ("distance effect"; Figure 5D, left). Also for number symbols, the numerical values could be classified significantly above chance level but with lower accuracy (38.8% ± 2.9%; Figure 5C, right). Misclassifications hardly varied

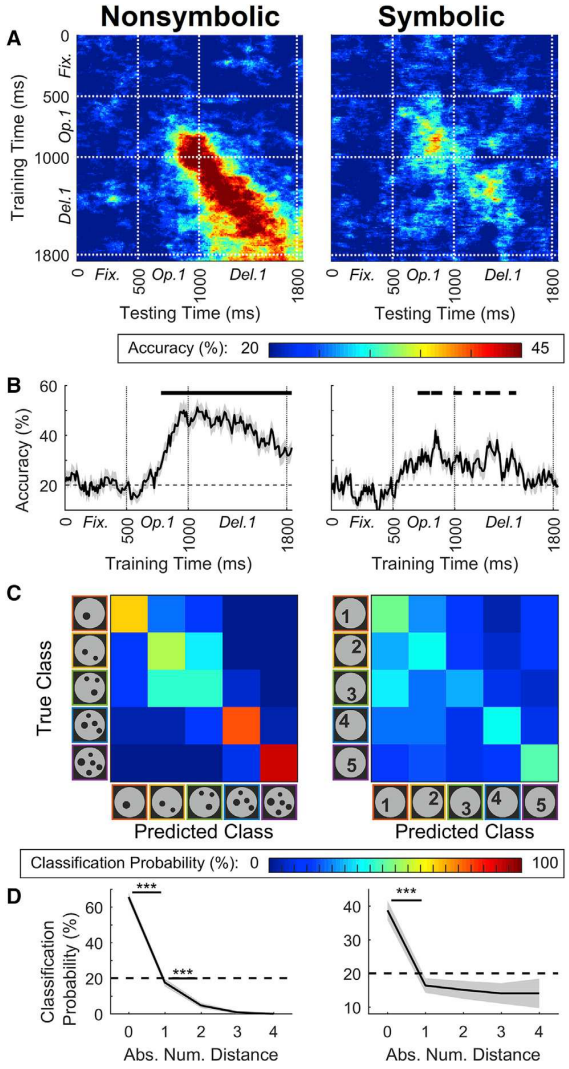


Figure 5. Numerosity Decoding Using a SVM Classifier

(A) Classification accuracy for decoding numerosity information when training a multi-class support vector machine (SVM) on instantaneous firing rates at a given time point and testing on another one for numerosity-selective (left) and numeral-selective (right) neurons.

(B) Accuracy for training and testing on identical time periods (main diagonal of matrices in A). The dashed line represents chance level (20% for five classes). Black bars above the data indicate significance ($p < 0.01$) when testing against performance for SVMs trained on shuffled data in a permutation test. Shaded areas indicate SEM.

(C) Confusion matrix derived when training an SVM on firing rates, averaged across the significant time windows in the temporal cross-training classification (B). Values on the main diagonal represent correct classification.

(D) Classification probability as a function of numerical distance. The dashed line represents chance level; shaded areas indicate SEM; asterisks represent significant differences between adjacent numerical distances ***$p < 0.001$).

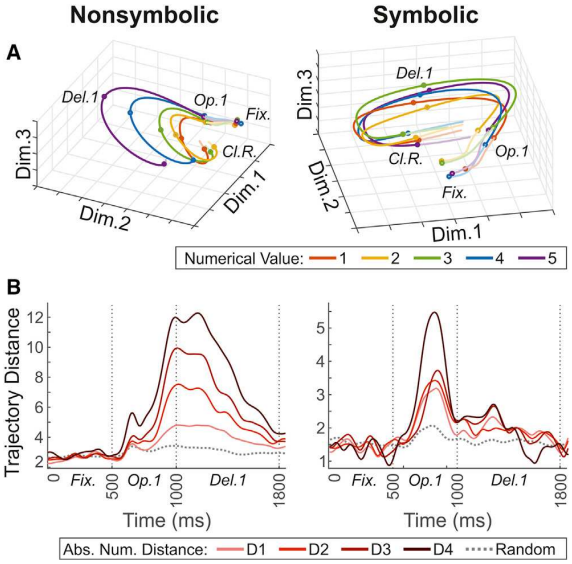


Figure 6. Population Dynamics based on State-Space Analysis

(A) Average state-space trajectories, reduced to the three principal dimensions for visualization, for the sub-populations of numerosity-selective (left) and numeral-selective (right) units. Each trajectory depicts the temporal evolution in the time window 0–1,850 ms. Circles indicate boundaries between experimental periods (*Cl.R.*, calculation rule; *Del.1*, delay 1; *Fix.*, fixation; *Op.1*, operand 1).

(B) Intertrajectory distances, averaged across pairs of trajectories with the same numerical distance. Dashed lines represent the average distances for trajectories obtained for label-shuffled data.

as a function of numerical distance for number symbols (Figure 5D, right). At absolute numerical distances 2, 3, and 4, classification probabilities obtained for the classifiers (n = 32) trained on nonsymbolic and symbolic number neurons were almost identical and significantly higher for symbolic than for nonsymbolic number ($p < 0.01$; t test). This indicates a sharper transition from the preferred to all nonpreferred numbers and thus greater selectivity in neurons tuned to symbolic number. When applied to the entire set of single units regardless of numerosity selectivity (585 units), this analysis yielded qualitatively similar results (Figure S4).

In addition, we analyzed the coding capacity and dynamics of the population of number-selective neurons by performing a multi-dimensional state-space analysis (see Supplemental Information) for nonsymbolic and symbolic numbers separately. At each point in time, the activity of *n* recorded neurons is defined by a point in *n*-dimensional space, with each dimension representing the activity of a single neuron. This results in trajectories that are traversed for different neuronal states, i.e., the five different numerical values in the nonsymbolic (Figure 6A, left) and symbolic format (Figure 6A, right). These trajectories reflect the instantaneous firing rates of the respective neuronal population as they evolve over time. To evaluate the temporal evolution of population numerical tuning in each format, we measured Euclidian distances between trial trajectories in the whole

population space corresponding to the activity to the five numerical values. In the nonsymbolic format, the trajectory distances systematically increased with numerical distance ($p < 0.001$; permutation test for all trajectories; see Supplemental Information), starting shortly after onset of operand 1 until the end of the memory delay 1. The distances between the population trajectories confirm the findings based on single selective neurons: the closer two numerosities were in the numerical continuum, the more similar were the corresponding patterns of population activity and vice versa (Figure 6B, left). This argues for a numerical distance effect in the population data. In the symbolic format, the trajectory distances were much less pronounced but likewise tended to increase with numerical distance ($p < 0.001$ for 1 versus 4 in a permutation test), reflecting the remnants of a distance effect (Figure 6B, right). A comparison of the trajectory distances also suggests that MTL neurons responded longer lasting to the nonsymbolic format and throughout the working memory period (i.e., delay 1). In contrast, the responses to the symbolic format were more confined to the sample phase of operand 1. Again, this analysis yielded similar results when performed for the whole population of single units (Figure S5).

Encoding of Number in Later Task Phases

After analysis of the responses to the operand 1, we also examined selectivity to the numerical value of operand 2 separately for nonsymbolic and symbolic number format. For the nonsymbolic format, 7.7% (45/585) of the tested neurons showed activity that varied exclusively with the number of operand 2 items during operand 2 presentation, irrespective of the dot array layout (5-factor sliding-window ANOVA with the factors numerical value of operand 1 [1–5], numerical value of operand 2 [0–5], protocol [standard and control], "mathematical rule" [addition and subtraction], and "rule cue" [word and symbol]; $\alpha = 0.01$). Twenty-two of the units selective to nonsymbolic operand 1 ($n = 92$) were also tuned to nonsymbolic operand 2; of those, 9 cells had the same preferred number. Given that 20% of the selective units are expected to share the preferred number by chance (5 number values), this proportion of 9 cells was significantly higher ($p < 0.05$ in binomial test). The finding that cells that responded both to operand 1 and operand 2 tended to show the same preferred numerosity was also confirmed by a correlation analysis (Pearson's $r = 0.64$; $p = 0.0013$; Figure S6). For the symbolic format, only a chance proportion of 1.5% (9/585) responded exclusively to the numerical value of the operand 2 during the presentation of the operand 2.

The responses of a single neuron throughout the whole trial are shown in Figure S7. This neuron was significantly tuned to numerosity 5 of operand 1 during the operand 1 phase and of operand 2 during the operand 2 phase (Figure S7, upper histograms). This neuron also showed strong responses to the numerical values of the operand 2 during the symbolic format (Figure S7, lower histograms); however, it was also selective to the numeral protocol and thus not counted as an exclusively numeral-selective cell. Overall, the highest proportion of neurons selective to the nonsymbolic numerical value of operand 2 in the MTL was found in the parahippocampal cortex (20%), followed by the hippocampus (6%; Figure S8).

Next, we analyzed selectivity to number in the delay 2 phase, again separately for nonsymbolic and symbolic number format. In the delay 2 phase, all the information necessary to solve the calculation is available to the subjects. The delay 2 phase may therefore be regarded as the calculation result phase. For statistical analysis, we applied a sliding-window 6-factor ANOVA (with the same factors as above, plus main factor numerical value corresponding to the result of the calculation [0–9]; $\alpha = 0.01$). Neither for the nonsymbolic nor for the symbolic format was the proportion of neurons selective to the calculation result higher than expected by chance (Figure S8).

Representation of Calculation Rules

Finally, we explored whether MTL neurons also encoded the calculation rules (addition and subtraction) in an abstract manner, independent from the rule notation (word or calculation symbol as rule cues). Cells selective to nonsymbolic numerical rules have been found in monkey cortex (Vallentin et al., 2012; Eiselt and Nieder, 2013). We determined calculation rule-selective units by applying a sliding-window 4-factor ANOVA (with the factors mathematical rule [addition and subtraction], rule cue [word and symbol], numerical value of operand 1 [1–5], and format [symbolic and nonsymbolic]; $\alpha = 0.01$) during the calculation rule phase and the rule delay phase. Figure S9 displays two rule-selective neurons. The neuron in Figure S9A showed a selective increase whenever an addition was required (reddish discharges), whereas the neuron in Figure S9B selectively enhanced discharges whenever a subtraction was cued (blueish colors). These rule-selective response increases were abstract and independent from the notation of the rule cue (word or symbol). In total, we found only a small proportion of 2% of abstract calculation rule cells, but this fraction was significantly larger than expected by chance (Figure S10). In addition, a significant fraction of 3% of the cells encoded the rule cue (calculation word or symbol) during the calculation rule phase (Figure S10).

DISCUSSION

Using single-cell recordings in subjects performing a calculation task, we have shown that single neurons in the MTL of humans are tuned to numerical values in nonsymbolic dot displays and symbolic numerals. The data about nonsymbolic number coding from humans can now be compared to those of nonhuman primates. In addition, our MTL recordings show how the capacity to represent symbolic number is represented in this part of our brain. This capacity to link number to visual signs has precursors in nonhuman primates (Diester and Nieder, 2010; Livingstone et al., 2014), but ultimately the symbolic number system is uniquely human (Nieder, 2009).

Functional imaging studies in humans found that areas of the MTL—among many other functions outside of the number domain—participate in learning arithmetic (De Smedt et al., 2011; Supekar et al., 2013; Qin et al., 2014; Menon, 2016). Using single-cell recordings in human subjects, we show that MTL neurons encode the numerical values in both nonsymbolic and symbolic number. With 29% and 6% of all neurons being selective to nonsymbolic and symbolic number, respectively, the parahippocampal cortex (PHC) shows the highest proportions of number

neurons among the four tested MTL areas. The PHC is part of a large network that connects regions of the temporal, parietal, and frontal cortices and has been associated with many cognitive processes (Aminoff et al., 2013), such as selectivity to pictures (Kreiman et al., 2000), responses guided by familiarity (Rutishauser et al., 2006), responses to spatial factors (Jacobs et al., 2013), and responses to mirror actions (Mukamel et al., 2010). Most likely, representations about numerical quantity do not originate within the PHC (or other areas of the MTL) but are provided via direct anatomical connections to the parieto-frontal core number system (Goldman-Rakic et al., 1984). Interestingly, the PHC has prominent connections with polymodal association areas, including the parietal lobule (Suzuki, 2009). This connection with the parietal lobule, an integral part of the core number network (Piazza et al., 2004, 2007; Arsalidou and Taylor, 2011) in which numerosity, but not number symbols, are mapped topographically (Harvey et al., 2013) is likely to provide the PHC with semantic information about numerical magnitude.

We have discovered two largely segregated populations of tuned number neurons in the human MTL that process either nonsymbolic or symbolic numerical quantity. The representation of nonsymbolic and symbolic number information by two distinct populations of tuned number neurons may either be inherited from the core number system or a special feature of the human MTL. Neurons in the prefrontal cortex of monkeys have been shown to respond abstractly by integrating visual and auditory numerosity (Nieder, 2012). Of course, number neurons in nonhuman primates operate strictly within the nonsymbolic format, but in monkeys trained to associate visual shapes with varying numbers of items, the responses of prefrontal neurons to the visual shapes reflected the associated numerical value in a behaviorally relevant way (Diester and Nieder, 2007).

Irrespective of its neurophysiological realization, format dependency does not pose a conceptual problem to number coding. In the human functional imaging literature, it is debated to what extent neural representations of number even in the human intraparietal sulcus (IPS) are format independent (Piazza et al., 2007; Eger et al., 2009; Jacob and Nieder, 2009b; Damarla et al., 2016) or format dependent (Cohen Kadosh et al., 2007; Holloway et al., 2010). There is not even consensus with regard to the degree of abstractness of numerical representations (reviewed in Cohen Kadosh and Walsh, 2009). Of course, these findings derived from blood-oxygen-level-dependent signals might also be explained by functionally segregated circuits that overlap at the macroscopic voxel scale. Future single-cell recordings in human subjects, in particular in the parietal and frontal association cortices, may help to resolve the question of abstract or segregated number neurons. They could also provide insights into the coding of larger numbers, the empty set, and the special number zero (Merten and Nieder, 2012; Ramirez-Cardenas et al., 2016).

Our study also helps to answer the question of the neuronal code for number. Two competing hypotheses have been proposed. Numbers could either be encoded by a "summation code," as evidenced by monotonic discharges as a function of quantity (Roitman et al., 2007), or by a "labeled-line code" as witnessed by numerosity-selective neurons tuned to preferred numerosities. In agreement with influential computational models of number processing (Dehaene and Changeux, 1993; Verguts and Fias, 2004), the number neurons we found in the human MTL were tuned to their individual preferred numerical value. A general concern of data from patients with a history of epileptic seizures is of course that the functional properties of MTL neurons may have affected during the course of the disease. Moreover, eye movement that was not measured during human recordings might have influenced the neurons' response properties. However, such factors are unlikely responsible for our results, because the same code that we observed in MTL neurons has been found multiple times in single-cell recordings of fixating monkeys, both in trained (Nieder et al., 2002, 2006; Sawamura et al., 2002; Nieder and Miller, 2004; Nieder, 2012) and numerically naive subjects (Viswanathan and Nieder, 2013) and even in corvid birds (Ditz and Nieder, 2015). This coding similarity suggests that our findings in the MTL are representative also for the healthy human brain. In addition, it indicates that number coding in humans and other animals is best captured by a labeled-line code. Of course, because number neurons only represent a very restricted part of the number line, only populations of number neurons, each tuned to different values, can represent the entire "mental number line."

In order to link number neurons to numerical behavior, neuronal responses need to explain number judgments (Nieder and Miller, 2003; Pinel et al., 2004). The direct comparison of responses in error trials versus correct trials, an analysis regularly done in nonhuman primates, would have been informative, but the human subjects hardly made any error and thus precluded the evaluation of error trials. However, as a basic requirement supporting the link between neurons and behavior, we show that nonsymbolic and symbolic numerical values can be reliably decoded from MTL neurons (Ramirez-Cardenas et al., 2016). This holds true for the populations of selective number neurons but also for the entire population of recorded neurons and irrespective of response selectivity. In addition, the neuronal activity can also explain the numerical distance effect, the finding that numerically distant numbers can be better discriminated. Behavioral studies and neural modeling show that the distance effect is substantial for the comparison of nonsymbolic numerosities but minute for judgments of exact number symbols (Buckley and Gillman, 1974; Verguts and Fias, 2004). In agreement with this, the accuracy of number discrimination based on the neuronal discharges exhibited large distance effects for the populations of broadly tuned numerosity-selective neurons but small distance effects for sharply tuned numeral-selective neurons. This finding provides further evidence for these neurons as the physiological correlate of number representations.

The distance effect for number symbols is thought to be inherited from more basic nonsymbolic number representations (Moyer and Landauer, 1967; Buckley and Gillman, 1974; Piazza et al., 2007). Its presence in human number neurons therefore supports the hypothesis that high-level human numerical abilities are rooted in biologically determined mechanisms. It suggests that number symbols acquire their numerical meaning by becoming linked to evolutionarily conserved set size representations during cognitive development (Halberda et al., 2008; Szkudlarek and Brannon, 2017). Symbolic number cognition thus appears to be grounded in neuronal circuits devoted to deriving

precise numerical values from approximate numerosity representations (Dehaene and Cohen, 2007).

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Neurophysiological Recording
  - Stimuli
  - Experimental Task
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Sliding-Window 2-Factor Analysis of Variance (ANOVA)
  - Tuning Properties
  - Multi-Class Support Vector Machine (SVM) Classification
  - Population State-Space Analysis
  - Other Task Phases
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes ten figures and one table and can be found with this article online at https://doi.org/10.1016/j.neuron.2018.08.036.

ACKNOWLEDGMENTS

We thank all patients for their participation. This research was supported by the Volkswagen Foundation, Germany, and the German Research Council (MO930/4-1 and SFB1089).

AUTHOR CONTRIBUTIONS

A.N. and F.M. designed the study; C.E.E. and F.M. recruited patients; J.B. and F.M. implanted the electrodes; E.F.K. and F.M. collected the data; E.F.K. analyzed the data with contributions from A.N. and F.M.; and A.N., E.F.K., and F.M. wrote the paper. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 9, 2018
Revised: June 19, 2018
Accepted: August 24, 2018
Published: September 20, 2018

REFERENCES

Aminoff, E.M., Kveraga, K., and Bar, M. (2013). The role of the parahippocampal cortex in cognition. Trends Cogn. Sci. *17*, 379–390.

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

Brannon, E.M., and Terrace, H.S. (1998). Ordering of the numerosities 1 to 9 by monkeys. Science *282*, 746–749.

Buckley, P.B., and Gillman, C.B. (1974). Comparisons of digits and dot patterns. J. Exp. Psychol. *103*, 1131–1136.

Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 27.

Cohen Kadosh, R., and Walsh, V. (2009). Numerical representation in the parietal lobes: abstract or not abstract? Behav. Brain Sci. *32*, 313–328, discussion 328–373.

Cohen Kadosh, R., Cohen Kadosh, K., Kaas, A., Henik, A., and Goebel, R. (2007). Notation-dependent and -independent representations of numbers in the parietal lobes. Neuron *53*, 307–314.

Daitch, A.L., Foster, B.L., Schrouff, J., Rangarajan, V., Kaşikçi, I., Gattas, S., and Parvizi, J. (2016). Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. Proc. Natl. Acad. Sci. USA *113*, E7277–E7286.

Damarla, S.R., Cherkassky, V.L., and Just, M.A. (2016). Modality-independent representations of small quantities based on brain activation patterns. Hum. Brain Mapp. *37*, 1296–1307.

De Smedt, B., Holloway, I.D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. Neuroimage *57*, 771–781.

Dehaene, S. (1997). The Number Sense (Oxford: Oxford Univ. Press).

Dehaene, S., and Changeux, J.P. (1993). Development of elementary numerical abilities: a neuronal model. J. Cogn. Neurosci. *5*, 390–407.

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

Diester, I., and Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. PLoS Biol. *5*, e294.

Diester, I., and Nieder, A. (2010). Numerical values leave a semantic imprint on associated signs in monkeys. J. Cogn. Neurosci. *22*, 174–183.

Ditz, H.M., and Nieder, A. (2015). Neurons selective to the number of visual items in the corvid songbird endbrain. Proc. Natl. Acad. Sci. USA *112*, 7827–7832.

Eger, E., Michel, V., Thirion, B., Amadon, A., Dehaene, S., and Kleinschmidt, A. (2009). Deciphering cortical number coding from human brain activity patterns. Curr. Biol. *19*, 1608–1615.

Eiselt, A.K., and Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. J. Neurosci. *33*, 7526–7534.

Fried, I., MacDonald, K.A., and Wilson, C.L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron *18*, 753–765.

Gilmore, C.K., McCarthy, S.E., and Spelke, E.S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. Cognition *115*, 394–406.

Goldman-Rakic, P.S., Selemon, L.D., and Schwartz, M.L. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. Neuroscience *12*, 719–743.

Gordon, P. (2004). Numerical cognition without words: evidence from Amazonia. Science *306*, 496–499.

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

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

Holloway, I.D., Price, G.R., and Ansari, D. (2010). Common and segregated neural pathways for the processing of symbolic and nonsymbolic numerical magnitude: an fMRI study. Neuroimage *49*, 1006–1017.

Jacob, S.N., and Nieder, A. (2009a). Tuning to non-symbolic proportions in the human frontoparietal cortex. Eur. J. Neurosci. *30*, 1432–1442.

Jacob, S.N., and Nieder, A. (2009b). Notation-independent representation of fractions in the human parietal cortex. J. Neurosci. *29*, 4652–4657.

Jacobs, J., Weidemann, C.T., Miller, J.F., Solway, A., Burke, J.F., Wei, X.X., Suthana, N., Sperling, M.R., Sharan, A.D., Fried, I., and Kahana, M.J. (2013). Direct recordings of grid-like neuronal activity in human spatial navigation. Nat. Neurosci. *16*, 1188–1190.

Kreiman, G., Koch, C., and Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. Nat. Neurosci. *3*, 946–953.

Livingstone, M.S., Pettine, W.W., Srihasam, K., Moore, B., Morocz, I.A., and Lee, D. (2014). Symbol addition by monkeys provides evidence for normalized quantity coding. Proc. Natl. Acad. Sci. USA *111*, 6822–6827.

Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.

Menon, V. (2016). Memory and cognitive control circuits in mathematical cognition and learning. Prog. Brain Res. *227*, 159–186.

Merten, K., and Nieder, A. (2009). Compressed scaling of abstract numerosity representations in adult humans and monkeys. J. Cogn. Neurosci. *21*, 333–346.

Merten, K., and Nieder, A. (2012). Active encoding of decisions about stimulus absence in primate prefrontal cortex neurons. Proc. Natl. Acad. Sci. USA *109*, 6289–6294.

Mormann, F., Dubois, J., Kornblith, S., Milosavljevic, M., Cerf, M., Ison, M., Tsuchiya, N., Kraskov, A., Quiroga, R.Q., Adolphs, R., et al. (2011). A category-specific response to animals in the right human amygdala. Nat. Neurosci. *14*, 1247–1249.

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

Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. Curr. Biol. *20*, 750–756.

Nieder, A. (2009). Prefrontal cortex and the evolution of symbolic reference. Curr. Opin. Neurobiol. *19*, 99–108.

Nieder, A. (2012). Supramodal numerosity selectivity of neurons in primate prefrontal and posterior parietal cortices. Proc. Natl. Acad. Sci. USA *109*, 11860–11865.

Nieder, A. (2016). The neuronal code for number. Nat. Rev. Neurosci. *17*, 366–382.

Nieder, A., and Merten, K. (2007). A labeled-line code for small and large numerosities in the monkey prefrontal cortex. J. Neurosci. *27*, 5986–5993.

Nieder, A., and Miller, E.K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. Neuron *37*, 149–157.

Nieder, A., and Miller, E.K. (2004). A parieto-frontal network for visual numerical information in the monkey. Proc. Natl. Acad. Sci. USA *101*, 7457–7462.

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

Nieder, A., Diester, I., and Tudusciuc, O. (2006). Temporal and spatial enumeration processes in the primate parietal cortex. Science *313*, 1431–1435.

Niediek, J., Boström, J., Elger, C.E., and Mormann, F. (2016). Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours. PLoS ONE *11*, e0166598.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. Neuron *44*, 547–555.

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

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

Pinel, P., Piazza, M., Le Bihan, D., and Dehaene, S. (2004). Distributed and overlapping cerebral representations of number, size, and luminance during comparative judgments. Neuron *41*, 983–993.

Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D.C., and Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. Nat. Neurosci. *17*, 1263–1269.

Quiroga, R.Q., Reddy, L., Kreiman, G., Koch, C., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. Nature *435*, 1102–1107.

Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. Curr. Biol. *26*, 1285–1294.

Reber, T.P., Faber, J., Niediek, J., Boström, J., Elger, C.E., and Mormann, F. (2017). Single-neuron correlates of conscious perception in the human medial temporal lobe. Curr. Biol. *27*, 2991–2998.e2.

Roitman, J.D., Brannon, E.M., and Platt, M.L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. PLoS Biol. *5*, e208.

Rutishauser, U., Mamelak, A.N., and Schuman, E.M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron *49*, 805–813.

Sawamura, H., Shima, K., and Tanji, J. (2002). Numerical representation for action in the parietal cortex of the monkey. Nature *415*, 918–922.

Scarf, D., Hayne, H., and Colombo, M. (2011). Pigeons on par with primates in numerical competence. Science *334*, 1664.

Starr, A., Libertus, M.E., and Brannon, E.M. (2013). Number sense in infancy predicts mathematical abilities in childhood. Proc. Natl. Acad. Sci. USA *110*, 18116–18120.

Supekar, K., Swigart, A.G., Tenison, C., Jolles, D.D., Rosenberg-Lee, M., Fuchs, L., and Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. Proc. Natl. Acad. Sci. USA *110*, 8230–8235.

Suzuki, W.A. (2009). Comparative analysis of the cortical afferents, intrinsic projections and interconnections of the parahippocampal region in monkeys and rats. In The Cognitive Neurosciences IV, M.S. Gazzaniga, ed. (Cambridge: MIT Press), pp. 659–674.

Szkudlarek, E., and Brannon, E.M. (2017). Does the approximate number system serve as a foundation for symbolic mathematics? Lang. Learn. Dev. *13*, 171–190.

Vallentin, D., Bongard, S., and Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. J. Neurosci. *32*, 6621–6630.

Verguts, T., and Fias, W. (2004). Representation of number in animals and humans: a neural model. J. Cogn. Neurosci. *16*, 1493–1504.

Viswanathan, P., and Nieder, A. (2013). Neuronal correlates of a visual "sense of number" in primate parietal and prefrontal cortices. Proc. Natl. Acad. Sci. USA *110*, 11187–11192.

Wynn, K. (1992). Addition and subtraction by human infants. Nature *358*, 749–750.

Xu, F., and Spelke, E.S. (2000). Large number discrimination in 6-month-old infants. Cognition *74*, B1–B11.

Yu, B.M., Cunningham, J.P., Santhanam, G., Ryu, S.I., Shenoy, K.V., and Sahani, M. (2009). Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. J. Neurophysiol. *102*, 614–635.

Please cite this article in press as: Kutter et al., Single Neurons in the Human Brain Encode Numbers, Neuron (2018), https://doi.org/10.1016/j.neuron.2018.08.036

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Cheetah software	Neuralynx Inc.	https://neuralynx.com/software/cheetah
Combinato spike sorting software	Niediek et al. (2016)	https://github.com/jniediek/combinato
MATLAB R2017a	MathWorks	https://de.mathworks.com/
Psychtoolbox		http://psychtoolbox.org/
LIBSVM	Chang and Lin (2011)	https://www.csie.ntu.edu.tw/~cjlin/libsvm/
DataHigh	Yu et al. (2009)	https://users.ece.cmu.edu/~byronyu/software/DataHigh/datahigh.html
Other		
Behnke-Fried depth electrodes	AD-TECH Medical Instrument Corp.	https://adtechmedical.com/depth-electrodes
ATLAS neurophysiology system	Neuralynx Inc.	https://neuralynx.com/news/techtips/atlas-neurophysiology-system-for-cogneuro-applications

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources should be directed to and will be fulfilled by Florian Mormann (florian.mormann@ukbonn.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Nine human subjects (4 male, all right-handed, mean age 43.3 years) undergoing treatment for pharmacologically intractable epilepsy participated in the study. Informed written consent was obtained from each patient. All studies conformed to the guidelines of the Medical Institutional Review Board at the University of Bonn, Germany. On the level of single neurons no sex- or gender-specific differences are to be expected; thus, the influence of sex and gender identity was not analyzed further.

METHOD DETAILS

Neurophysiological Recording

All subjects were implanted bilaterally with chronic intracerebral depth electrodes in the medial temporal lobe (MTL) to localize the epileptic focus for possible clinical resection. The exact electrode numbers and locations varied across subjects and were based exclusively on clinical criteria. Neuronal signals were recorded using 9–10 clinical Behnke-Fried depth electrodes (AD-TECH Medical Instrument Corp., Racine, WI). Each electrode contained a bundle of nine platinum-iridium micro-electrodes protruding from its tip; eight high-impedance active recording channels, and one low-impedance reference electrode. Differential neuronal signals (recording range ± 3200 μV) were filtered (bandwidth 0.1–9,000 Hz), amplified and digitized (sampling rate 32.7 kHz) using a 256-channel ATLAS neurophysiology system (Neuralynx Inc., Bozeman, MT). Behavioral data were synchronized with the recorded spikes via 8-bit timestamps using the Cheetah software (Neuralynx Inc., Bozeman, MT).

After band-pass filtering the signals (bandwidth 300–3,000 Hz), spikes were detected and pre-sorted automatically using the Combinato software (Niediek et al., 2016). Manual verification and classification as artifact, multi or single unit was based on spike shape and its variance, inter-spike interval distribution per cluster and the presence of a plausible refractory period. Only units that responded with an average firing rate of > 1 Hz during operand 1 and delay 1 phase for either format were included in the analyses. Across 16 recording sessions from all nine patients, a total of 836 units (585 single and 251 multi units) were identified in the amygdala (AMY; 153 single and 63 multi units), parahippocampal cortex (PHC; 126 single and 61 multi units), entorhinal cortex (EC; 107 single and 54 multi units) and hippocampus (HIPP; 199 single and 73 multi units) according to these criteria (see Table S1); 333 units with firing rates < 1 Hz were excluded. Only single units were subjected to further analyses.

Stimuli

All stimuli were presented within a filled gray circle (diameter 6° of visual angle) on a black background. During fixation and delay phases, a white fixation spot was presented in the center of the gray area. It disappeared during stimulus presentation to avoid

confusion with nonsymbolic stimuli and to distinguish it clearly from the nonsymbolic zero-stimulus that was included for control purposes as a potential operand 2-stimulus (see *Experimental Task*).

Number stimuli of operand 1 ranged from 1 to 5, and were either black 'symbolic' Arabic digits at a randomized location ('numerals'), or 'nonsymbolic' arrays of black dots of pseudo-randomly varied sizes and at randomized locations where the number of dots corresponded to the respective numerical value ('numerosities'). Number stimuli of operand 2 ranged from 0 to 5, and were the same as for operand 1. For the nonsymbolic 'zero'-stimulus the empty gray circle without fixation spot was presented.

Both nonsymbolic and symbolic number formats were shown in standard and control displays, or 'protocols' (Figure 1B). This was done in order to control for low-level visual features. The standard nonsymbolic numerosity displays consisted of randomly placed dots of varying sizes (diameter 0.3° to 0.8° of visual angle), whereas in the control displays the overall surface area and density of the dots across numerosities was equated. For the Arabic numerals, different font types were used as standard (Helvetica, 34 pt) and control (DS-Digital, 34 pt) displays. A session consisted of 50% nonsymbolic and 50% symbolic number formats. Within each format, standard and control protocols were shown with equal probability of 50%.

Two different mathematical rules, i.e., addition and subtraction, were applied (Figure 1C). To dissociate neuronal activity related purely to physical properties of the operator from the rule that it signifies, two distinct cues, i.e., the mathematical sign (+ or –) or a verbal analog ('und' [add] and 'weniger' [subtract]), were used for each rule (all Helvetica, 34 pt, and presented in the center).

Experimental Task

During experimental sessions, subjects sat in bed and performed the task on a touch-screen laptop (display diagonal 11.7 in; resolution 1366x768 px) on which stimuli were presented at a distance of approximately 50 cm. To exclude any bias, the subjects were not informed about the purpose or hypotheses of the experiment.

Subjects performed two calculation tasks that required them to calculate the result of a simple arithmetic problem (Figure 1A). Each trial started with a 500 ms fixation phase. Then, stimuli were presented successively in the order operand 1 – operator – operand 2 for 500 ms each, followed each by 800 ms delay phases. Afterward, a number pad showing the Arabic numerals 0 to 9 was presented on the screen and subjects were instructed to touch the number matching the result of the calculation in a self-paced manner. After a 500 ms feedback display ('richtig' [correct] or 'falsch' [false]) the next trial was started automatically.

We varied five factors in this task: Format (symbolic/ nonsymbolic), numerical value (1–5), and protocol (standard/ control) for the operand 1-stimulus, resulting in 20 different 'number' conditions, as well as mathematical rule (addition/ subtraction) and rule cue (symbol/ word), resulting in four 'operator' conditions. Operand 2 was always of the same format and protocol as operand 1, but with random numerical value 0–5, albeit guaranteeing calculation results between 0 and 9.

Each session comprised a total of 320 trials, plus 10 rehearsal trials at the beginning to familiarize subjects with the task that were excluded from further analysis. A session was divided into four blocks of 80 trials each, comprising each of the 80 different conditions in pseudo-random order, to allow for short self-paced breaks in between. Thus, every number condition (i.e., combination of number, format and protocol) was presented 16 times, while every operator condition (i.e., combination of rule and rule cue) occurred 80 times.

QUANTIFICATION AND STATISTICAL ANALYSIS

Only single units (n = 585) were included in the following analyses. We focused on the operand 1 and delay 1 phases because these were the only periods during which pure number information was being processed. Given that the rule to be applied was not yet known, interference of calculation processes or motor response preparation could be excluded. Thus, all analyses were conducted for the time window 0–1850 ms (fixation onset to delay 1 offset). All subjects performed the task with high proficiency (98.5% ± 0.6%, range 90.3%–99.8%). Therefore, we did not exclude the negligible number of error trials from the analyses.

Sliding-Window 2-Factor Analysis of Variance (ANOVA)

Due to the incomparability of the protocol conditions for the different formats, the following procedure was carried out separately for trials of each format. For each unit, spike trains were smoothed trial-wise (Gaussian kernel, $\sigma = 150$ ms) within the analysis window. At every 10-ms-step, a 2-factor ANOVA was performed on the instantaneous firing rates for the factors 'numerical value' (1–5) and 'protocol' (standard/ control) resulting in a temporal sequence of *F*-values for main and interaction effects. To control for multiple comparisons, a cluster permutation test (Maris and Oostenveld, 2007) was performed to identify temporal clusters that encoded number information significantly. Briefly, all *F*-values within a cluster, i.e., an interval with only significant *p*-values ($p_{clus} < 0.01$) for the respective effect, were summed up. Calculating multiple 2-factor ANOVAs and summing up significant *F*-values was repeated with randomly shuffled trial labels ($n_{perm} = 100$). A temporal cluster of the true data was then considered significant only if the percentile rank of the summed *F*-values of the true data was significant across the distribution of summed *F*-values obtained for the shuffled data ($p_{rank} < 1\%$, corresponding to a nominal size of the statistical test of $\alpha = 0.01$). In the following, we refer to such a significant cluster as NUM-interval. A unit was counted as exclusively number-selective ('number-unit') if a significant cluster was observed between 500–1600 ms (operand 1 onset to 200 ms before delay 1 offset) for the factor 'numerical value' and there were no overlapping significant clusters for the factor 'protocol' or the interaction (see Figure S1). As a control, we determined the proportion of significant NUM-intervals for the shuffled data (585 single units x 100 permutations, resulting in 58,500 tests; same procedure as for the true

data) in order to estimate the probability of false positives, i.e., the probability that a unit was classified as number-selective by chance. For both formats, we found that 1% of these tests (nonsymbolic 493/58,500, symbolic 513/58,500) resulted in a statistically significant result, or false positive, confirming the empirical size of the statistical test to also be at $\alpha \approx 0.01$. The probability that neuronal selectivity occurs by chance was therefore 1%. Using a binomial test with $p_{chance} = \alpha = 0.01$, we can thus confirm that the observed proportion of number-selective neurons cannot be explained by chance occurrences both for nonsymbolic (92/585; $p_{binomial}$ = 1.18e-77) and symbolic (16/585; $p_{binomial}$ = 3.58e-4) number-selective neurons.

To compare the general response behavior of number-units and non-selective cells, we determined the maximum firing rate per number condition for each format and cell by averaging the spike rates within the significant NUM-interval for the number-units or across the entire operand 1 and delay 1 phase (500–1800 ms) for the non-selective units (nonsymbolic format: 92 numerosity-cells, 493 non-selective cells; symbolic format: 16 numeral-cells, 569 non-selective cells). Distributions were then compared using a Mann-Whitney-U-test (see Figure S2). The correlation between nonsymbolic and symbolic number-representations was evaluated for the sub-populations of numerosity-selective neurons (n = 92), numeral-selective neurons (n = 16) and neurons responsive to both formats (n = 6). For each unit of a sub-population, we calculated the preferred number for the significant format by averaging the spike rates during the respective NUM-interval, and the preferred number for the non-significant format by averaging the spike rates across the entire operand 1 and delay 1 phase (500–1800 ms). We then quantified the relationship by calculating Pearson's linear correlation coefficient (see Figure S3).

The sub-populations of nonsymbolic and symbolic number-units obtained with the sliding-window 2-factor ANOVA showed little (although significant) overlap. Therefore, the following population analyses were performed separately for the sub-population of nonsymbolic number-units (92 units) considering nonsymbolic trials only, and the sub-population of neurons preferring symbolic stimuli (16 units) using symbolic trials only. For control purposes, population analyses were also performed for the whole population of single units (585 units; see Figures S4 and S5).

Tuning Properties

For each number-unit, individual tuning curves were obtained by averaging the responses to different numerical values across trials, during the time window of significant number-clusters (NUM-intervals). In cases where we identified multiple NUM-intervals within the same unit, tuning curves were calculated separately for each of these intervals (3/5 nonsymbolic number-cells with multiple NUM-intervals preferred different numerosities). They were then normalized by setting the maximum response to 100% and the minimum response to 0%. The preferred numerical value was determined as the number which elicited the strongest average response. A cross-validation analysis was performed to estimate the robustness and reliability of the preferred number assessment (Nieder and Merten, 2007). We split the data into two halves by randomly assigning the trials to either of the two sets and calculated the preferred number for each dataset. This was done for the entire population of number-units and the relationship between preferred numbers quantified by calculating Pearson's linear correlation coefficient. If both datasets resulted in identical preferred numbers, the correlation coefficient was 1. The correlation analysis was performed 100 times for different random partitions of the data, and the average correlation coefficient was calculated.

Population neural filter functions were then calculated by averaging across the sub-populations of units preferring the same numerical value. The activity of each number-unit was considered as a function of distance from its preferred number. Differences between all pairs of adjacent numerical distances were separately quantified using Wilcoxon signed-rank tests. Moreover, for each numerical distance we tested whether the obtained response differed significantly from a response pattern to be expected in case of random tuning (obtained by repeating the analysis with shuffled labels) using a permutation test (n_{perm} = 1000).

Multi-Class Support Vector Machine (SVM) Classification

For each unit, spike trains were trial-wise smoothed (Gaussian kernel, σ = 50 ms, window size 300 ms) within the analysis window. For temporal cross-training classification, a multi-class SVM classifier (Chang and Lin, 2011) was trained on the instantaneous firing rates at a certain time point, and then tested on firing rates at different time points (sampling interval 10 ms). We used a linear SVM-kernel with default parameter settings and applied 'one-versus-one' classification to distinguish our five classes. For the 32 trials per number and format (symbolic versus nonsymbolic), we used leave-one-out cross-validation and normalized all firing rates by z-scoring (mean and standard deviation obtained from training data only) within each cross-validation repetition. For each classifier (n = 32), accuracy was assessed by counting the instances that a certain activity pattern was labeled correctly. To evaluate whether accuracy differed significantly from chance level (20% for five classes) when trained and tested at the same time points, we repeated the analysis with randomly shuffled trial labels (n_{perm} = 1000) and applied a cluster permutation test (p_{clus} = 0.01, p_{rank} = 1%; see *Sliding-Window 2-Factor ANOVA*). Finally, a multi-class SVM (with the same settings as above) was trained and tested on the firing rates obtained by averaging across the time window that was significant in the cross-training classification, i.e., window 780–1800 ms for the nonsymbolic number-units and 810–1370 ms for the symbolic number-cells (in cases where we obtained multiple significant windows, we used the onset of the first cluster and the offset of the last cluster as window boundaries). In addition to the overall accuracy, we assembled a confusion matrix which counted the frequency at which a trial of a certain stimulus class was assigned different labels by the classifier (main diagonal indicating correct labeling), and calculated the classification probabilities per numerical distance by averaging over the main and minor diagonals of the confusion matrix for each classifier (n = 32). Differences between adjacent classification probabilities were evaluated using Wilcoxon signed-rank tests (n = 32).

Population State-Space Analysis

For each unit, spike trains were averaged across conditions, normalized by z-scoring and smoothed (Gaussian kernel, σ = 50 ms, window size 300 ms). The temporal evolution of the neural activity of a population of *n* neurons can be represented as a trajectory in an *n*-dimensional space where each axis represents the instantaneous firing rate of one neuron. In our case, we analyzed the trajectories of the five different number conditions in a 92-dimensional space for the sub-population of nonsymbolic number-units, and in a 16-dimensional space for the symbolic number-units, respectively. To evaluate population tuning in terms of numerical distances, we calculated the Euclidean distances between each pair of trajectories, and averaged across those with the same numerical distance. This analysis was repeated with shuffled trial labels (n_{perm} = 1000) to obtain intertrajectory distances that would be expected for random numerical tuning, and evaluated using a cluster permutation test (p_{clus} = 0.01, p_{rank} = 1%; see *Sliding-Window 2-Factor ANOVA*). Solely for visualization purposes, trajectories were reduced to the top 3 (in terms of covariance they explain) orthonormalized dimensions using a Gaussian-process factor analysis (Yu et al., 2009).

Other Task Phases

Number-selectivity to operand 2 was assessed by performing a sliding-window 5-factor ANOVA with the factors 'numerical value' of operand 1 (1–5), 'numerical value' of operand 2 (0–5), 'protocol' (standard/ control), 'mathematical rule' (addition/ subtraction) and 'rule cue' (word/ symbol) for the operand 2 phase (analysis window 3050–3650 ms), separately for each format. We used the same parameters and procedures as for the operand 1 phases (see *Sliding-Window 2-Factor ANOVA*). A unit was counted as exclusively number-selective during the operand 2 phase if a significant cluster was observed between 3100–3400 ms (operand 2 onset to 200 ms before operand 2 offset) for the factor 'numerical value' of operand 2 and there were no overlapping significant clusters for any other factor. For the population of nonsymbolic number-units responsive to both operand 1 and 2 (n = 22) we calculated the preferred number per operand during the respective significant NUM-interval and quantified the relationship by calculating Pearson's linear correlation coefficient (see Figure S6). In addition, the significance of the proportion of units preferring the same number (k = 9) was evaluated using a binomial test (p_{chance} = 0.2 for five numbers).

Analogously, number-selectivity to the calculation result was determined for the delay 2 phase (analysis window 3550–4450 ms; we excluded the actual response phase in order to avoid confounds with motor responses) using a 6-factor ANOVA with the same factors as above, plus 'numerical value' of calculation result (0–9). Again, we used the same parameters and procedures as for the operand 1 phases (see *Sliding-Window 2-Factor ANOVA*) and counted a unit as exclusively number-selective if a significant cluster was observed between 3600–4200 ms (delay 2 onset to 200 ms before delay 2 offset) for the factor 'numerical value' of calculation result and there were no overlapping significant clusters for any other factor.

Furthermore, we determined rule-selective units by calculating a sliding-window 4-factor ANOVA with the factors 'mathematical rule' (addition/ subtraction), 'rule cue' (word/ symbol), 'numerical value' of operand 1 (1–5) and 'format' (symbolic/ nonsymbolic), thereby pooling over the factor 'protocol' (given its irrelevance for the processing of the rule cues), for the calculation rule and rule delay phases (analysis window 1750–3150 ms). The same parameters and procedures as for the operand 1 phases (see *Sliding-Window 2-Factor ANOVA*) were used. A unit was counted as exclusively rule-selective if a significant cluster was observed for the factor 'mathematical rule' between 1800–2900 ms (calculation rule onset to 200 ms before rule delay offset) and there were no overlapping significant clusters for any other factor. Exclusive cue-selectivity was defined analogously.

DATA AND SOFTWARE AVAILABILITY

Data and analysis software for this paper are available from the lead contact upon reasonable request.

Neuron, Volume *100*

Supplemental Information

Single Neurons in the Human Brain Encode Numbers

Esther F. Kutter, Jan Bostroem, Christian E. Elger, Florian Mormann, and Andreas Nieder

Table S1: Related to Figure 2. Neuronal Selectivity of MTL Single Units Across Patients. Proportions of exclusively number-selective single units, i.e. units with a significant main effect for 'numerical value', but no concurrent main effect for 'protocol' or any interaction effect in a 2-factor ANOVA evaluated at $\alpha = 0.01$, separately for each format (nonsymbolic vs. symbolic).

Patient-ID	**Nonsymbolic**		**Symbolic**	
1	43 %	(3/7)	0 %	(0/7)
2	15 %	(9/63)	2 %	(1/63)
3	22 %	(18/83)	2 %	(1/83)
4	17 %	(1/6)	0 %	(0/6)
5	0 %	(0/18)	6 %	(1/18)
6	16 %	(7/44)	3 %	(1/44)
7	12 %	(14/119)	2 %	(2/119)
8	15 %	(23/163)	4 %	(5/163)
9	21 %	(17/82)	7 %	(5/82)
TOTAL	**16 %**	**(92/585)**	**3%**	**(16/585)**

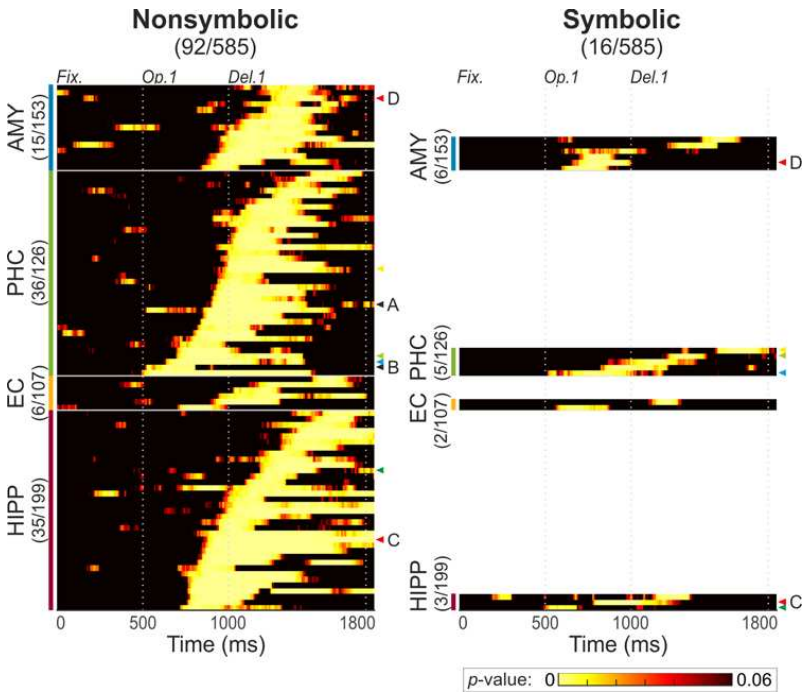


Figure S1: Related to Figure 2. Number-Encoding of MTL Single Units over Time. Temporal profiles of all exclusively number-selective units (proportions denoted in brackets), sorted by latency of number encoding, per MTL region (AMY–amygdala; PHC–parahippocampal cortex; EC–entorhinal cortex; HIPP–hippocampus). Each row shows the *p*-values of a unit for the factor 'numerical value' in the sliding-window 2-factor ANOVA, bright colours corresponding to significant NUM-intervals. The coloured triangles on the right of each panel indicate units that responded significantly to both numerosities and numerals; letters A–D refer to the example units shown in Figure 2.

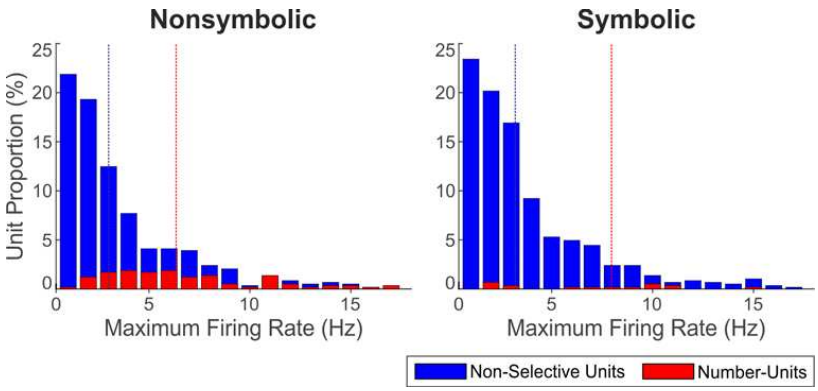


Figure S2: Related to Figure 2. Maximum Firing Rates of MTL Units. Distribution of maximum discharge rates for non-selective (blue) and number-selective (red) units in response to the presentation of nonsymbolic (left) and symbolic (right) number stimuli, respectively. Dotted lines represent median values. For both formats, number-selective units (median: nonsymbolic 6.3 Hz; symbolic 7.9 Hz) had significantly higher firing rates compared to non-selective units (median: nonsymbolic 3.0 Hz; symbolic 3.2 Hz) (nonsymbolic $p < 0.0001$; symbolic $p < 0.01$; Mann-Whitney-U-test). Discharge rates did not differ significantly between the sub-populations of numerosity- and numeral-selective units ($p = 0.73$).

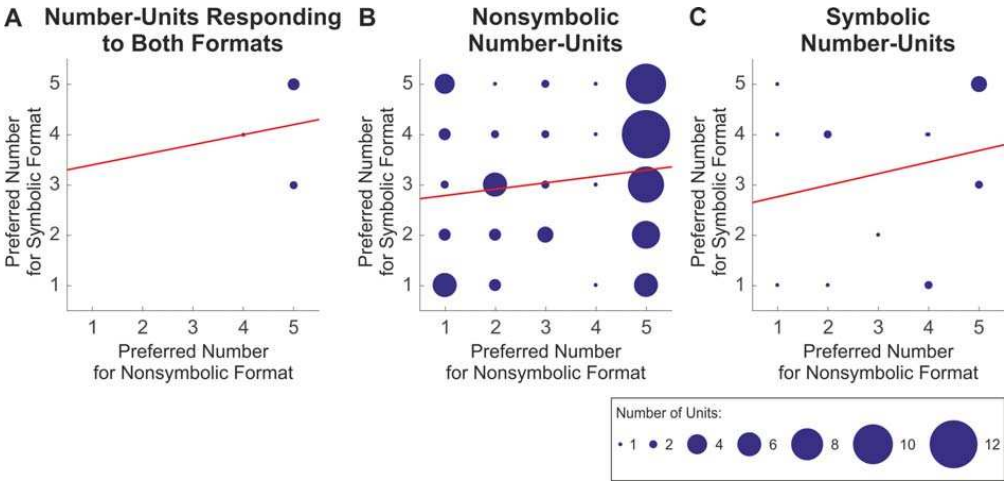


Figure S3. Related to Figure 2. Correlation between Nonsymbolic and Symbolic Number Representations. (A) Correlation of the preferred number of neurons ($n = 6$) significantly tuned to both nonsymbolic and symbolic number (Pearson's $r = 0.083$; $p = 0.87$). **(B)** Correlation of the preferred number of neurons significantly tuned to nonsymbolic numerosity ($n = 92$) with their (non-significant) tuning to symbolic numerals (Pearson's $r = 0.14$; $p = 0.19$). **(C)** Correlation of the preferred number of neurons significantly tuned to symbolic numerals ($n = 16$) with their (non-significant) tuning to nonsymbolic numerosity (Pearson's $r = 0.23$; $p = 0.39$).

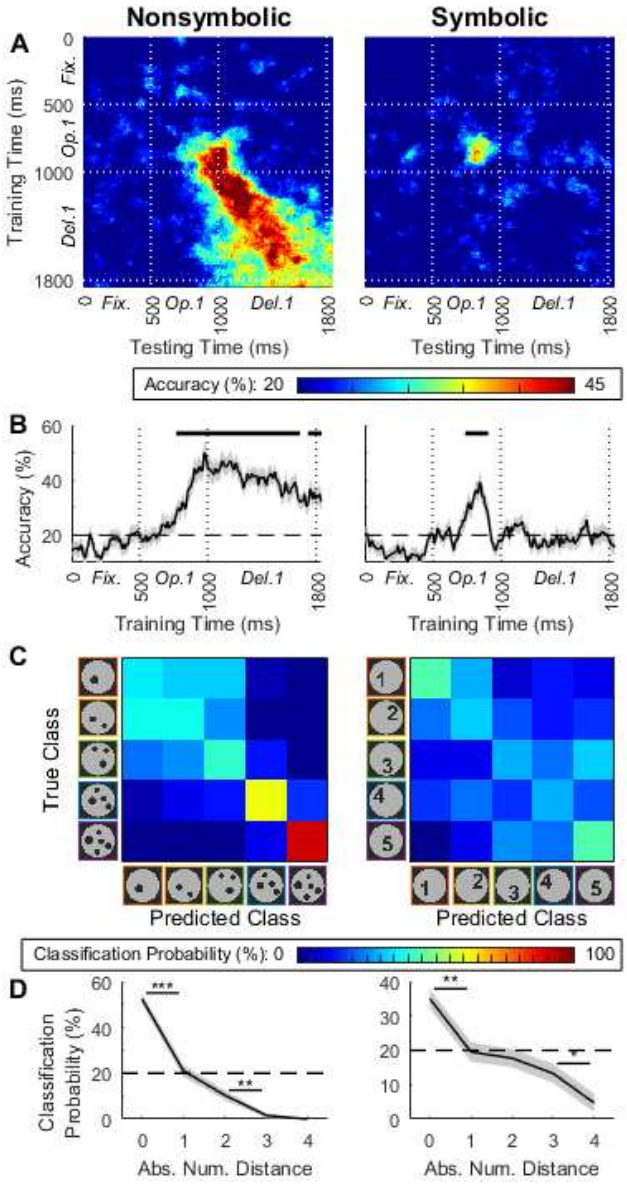


Figure S4: Related to Figure 5. Numerosity Decoding for the Whole Population of Single Units using a Support Vector Machine (SVM) Classifier. (A) Classification accuracy for decoding numerosity information when training a multi-class SVM on instantaneous firing rates at a given time point and testing on a different one. **(B)** Accuracy for training and testing on identical time periods (main diagonal of matrices in A). The dashed line represents chance level (20% for five classes); shaded areas indicate SEM. Black bars above the data indicate significance ($p < 0.01$) when testing against performance for SVMs trained on shuffled data in a permutation test. **(C)** Confusion matrix derived when training an SVM on firing rates, averaged across the significant time windows in the temporal cross-training classification (B). Values on the main diagonal represent correct classification. **(D)** Classification probability as a function of numerical distance. The dashed line represents chance level; shaded areas indicate SEM. Asterisks represent significant differences between adjacent numerical distances (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

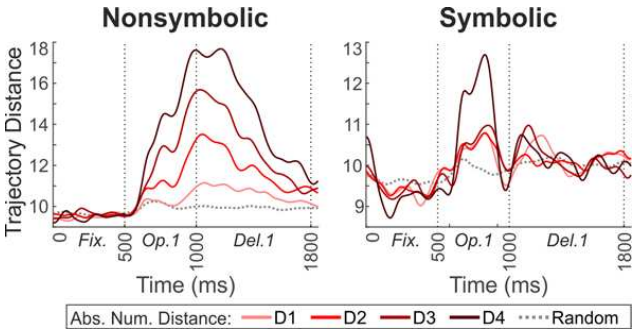


Figure S5: Related to Figure 6. Population Dynamics based on State-Space Analysis for the Whole Population of Single Units. Intertrajectory distances in 585-dimensional space, averaged across pairs of trajectories with the same numerical distance. Dashed lines represent the average distances for trajectories obtained for label-shuffled data. Vertical dotted lines indicate boundaries between experimental periods (*Fix.*–fixation; *Op.1*–operand 1; *Del.1*–delay 1).

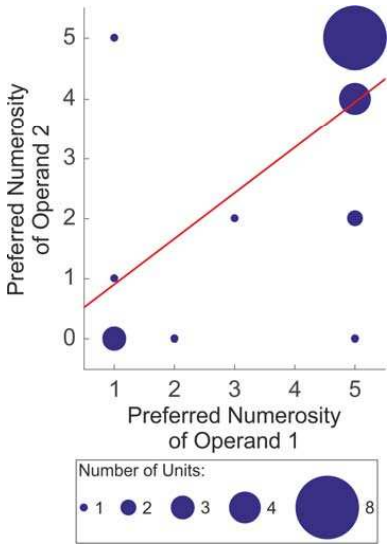


Figure S6. Related to Figure 2. Correlation between Preferred Numerosities during Operand 1 and Operand 2 Phases. Cells that responded to nonsymbolic number for both operand 1 and operand 2 (n = 22) tended to show the same preferred numerosity (Pearson's r = 0.64; p = 0.0013).

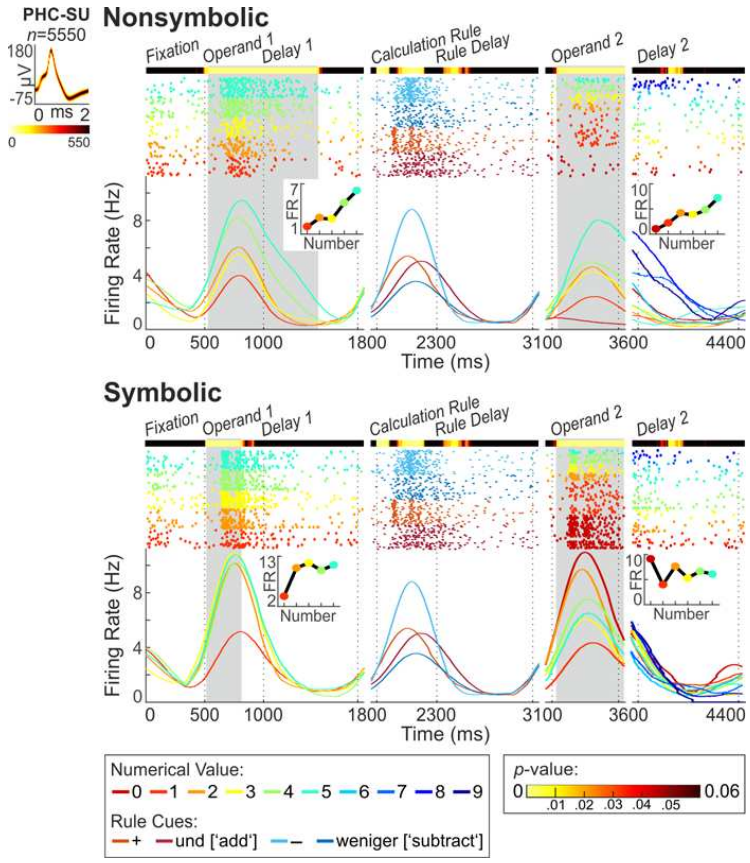


Figure S7: Related to Figure 2. Neuronal Responses of a Number-Selective Neuron during various Number-Related Task Phases. Responses to both nonsymbolic (upper row) and symbolic (lower row) number are shown. Neuronal activity is sorted according to the numerical value of operand 1 from fixation onset to the end of the operand 1 phase, according to rule cues during the calculation rule and rule delay phase, according to the numerical value of operand 2 during operand 2 phase, and according to the numerical value of the correct result of the calculation during the delay 2 phase. The left panel depicts a density plot of the recorded action potentials (colour darkness indicates number of overlapping wave forms according to color scale at the bottom). Panels show single-cell response rasters for many repetitions of the format (each dot represents an action potential) and averaged instantaneous firing rates below. Colours correspond to the different numerical values. Gray shaded areas represent significant number discrimination periods according to the respective sliding-window ANOVA (colour-coded *p*-values above each panel). Insets show the number tuning functions to operand 1 (1–5) and operand 2 (0–5), respectively.

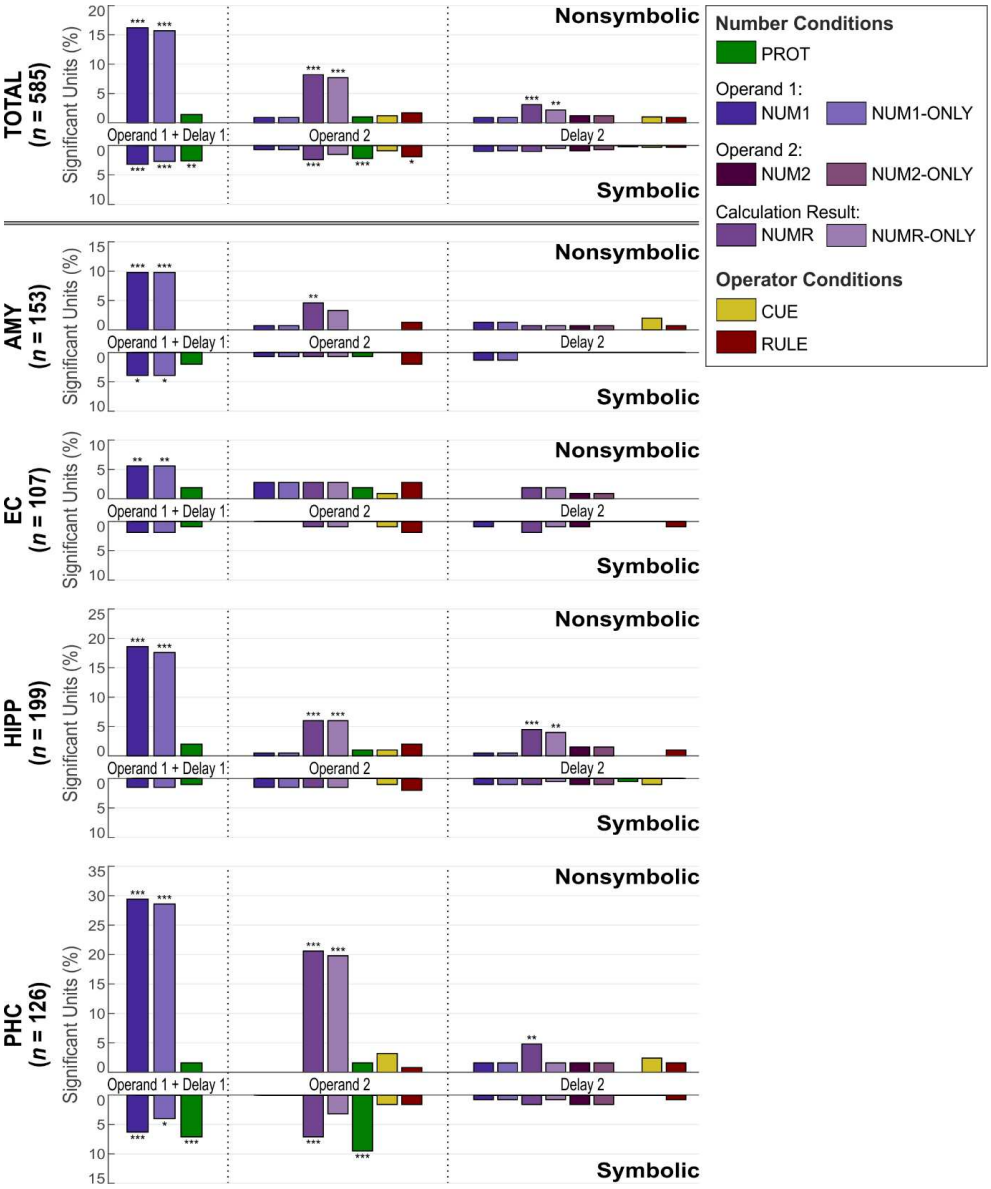


Figure S8: Related to Figure 3. Number Selectivity of MTL Single Units in all Number-Related Task Phases. Proportions of significant single units for different MTL regions (AMY–amygdala; EC–entorhinal cortex; HIPP–hippocampus; PHC–parahippocampal cortex) and different number-related task phases. Operand 1 phases: 2-factor ANOVA for 'protocol' (PROT: standard/ control) and 'numerical value' of operand 1 (NUM1: 1–5). Operand 2 phase: 5-factor ANOVA for 'protocol', 'numerical value' of operand 1 and operand 2 (NUM2: 0–5), 'mathematical rule' (RULE: addition/ subtraction) and 'rule cue' (CUE: symbol/ word). Calculation phase (delay 2): 6-factor ANOVA for 'protocol', 'numerical value' of operand 1, operand 2 and calculation result (NUMR: 0–9), 'mathematical rule' and 'rule cue'. Neurons with an effect for 'numerical value', but no concurrent other main effects are termed exclusively number-selective units (NUM-ONLY). All ANOVAs were evaluated at $\alpha = 0.01$, and numbers of significant neurons subjected to a Bonferroni-corrected ($n = 4$) binomial test; asterisks indicate significance ($^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$).

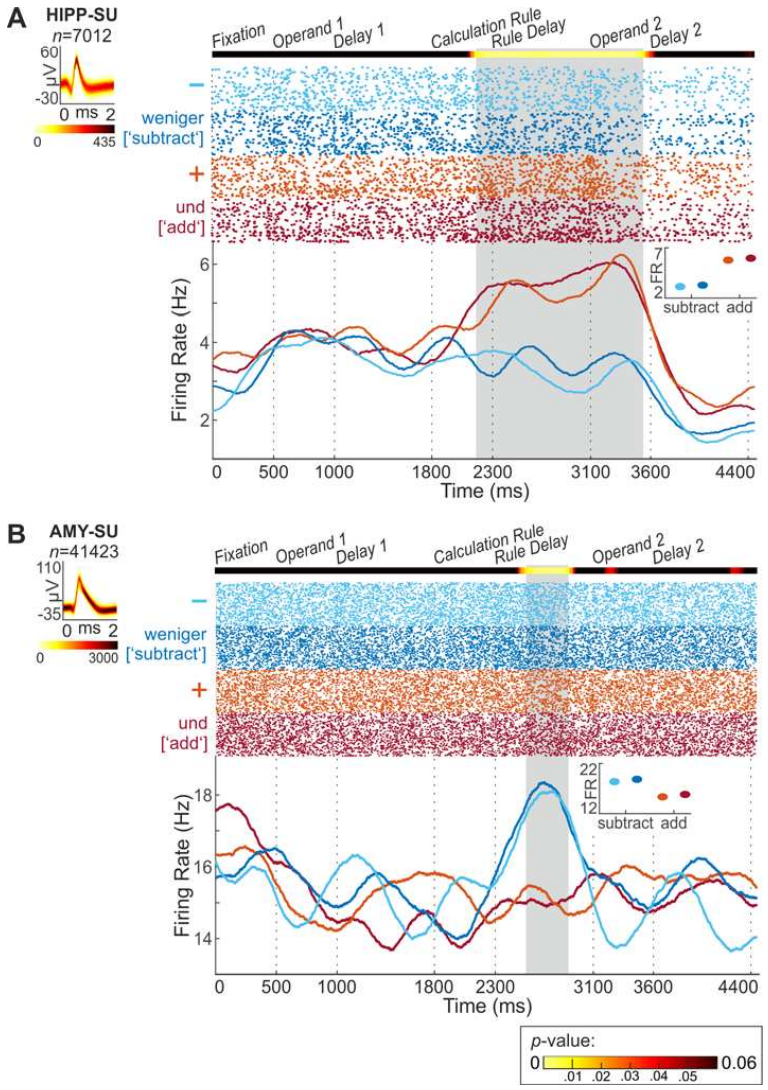


Figure S9: Related to Figure 2. Neuronal Responses of Rule-Selective Neurons. Responses of two example neurons selective for the 'addition' rule **(A)** and the 'subtraction' rule **(B)**. Left panels depict a density plot of the recorded action potentials (colour darkness indicates number of overlapping wave forms according to color scale at the bottom). Panels show single-cell response rasters for many repetitions of the rule cue (each dot represents an action potential) and averaged instantaneous firing rates below. Colours correspond to the four different rule cues. Insets show average activity per rule cue during the significant rule discrimination period according to the sliding-window ANOVA (colour-coded *p*-values above each panel).

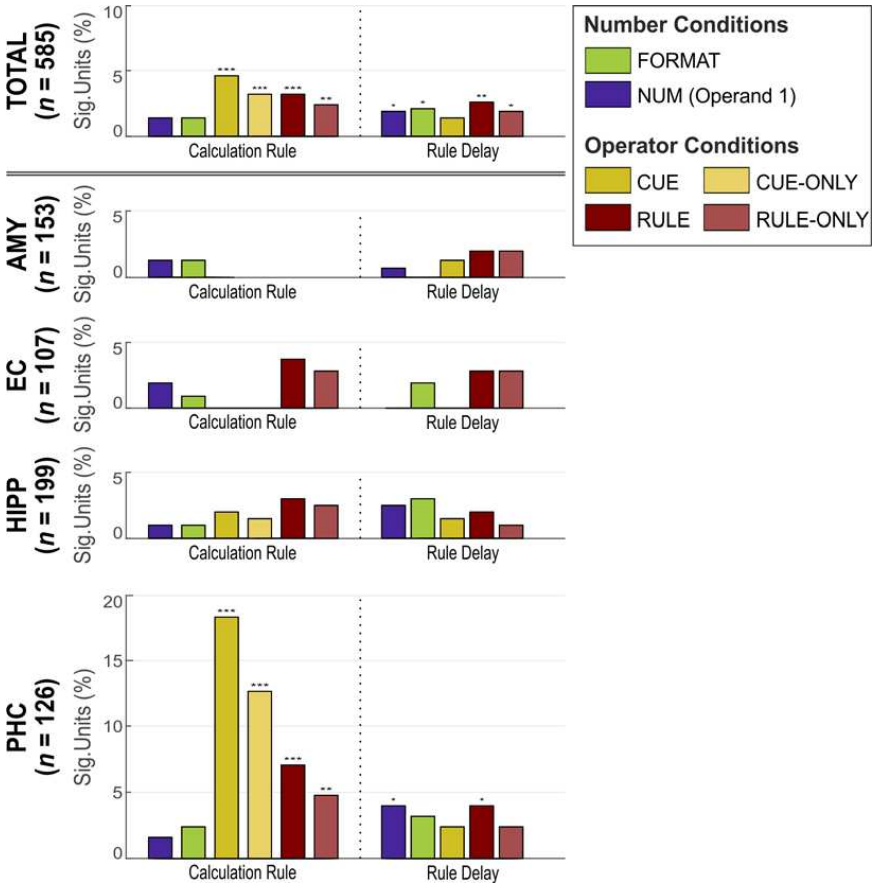


Figure S10: Related to Figure 3. Rule Selectivity of MTL Single Units. Proportions of significant single units with significant main effects for 'mathematical rule' (RULE: addition/ subtraction), 'rule cue' (word/ symbol), 'numerical value' of operand 1 (NUM: 1–5) and 'format' (FORMAT: symbolic/ nonsymbolic) in a 4-factor ANOVA evaluated at $\alpha = 0.01$, separately for each MTL region (AMY–amygdala; EC–entorhinal cortex; HIPP–hippocampus; PHC–parahippocampal cortex). Neurons with an effect for 'mathematical rule', but no concurrent other main effects are termed exclusively rule-selective units (RULE-ONLY), analogous for factor 'rule cue' (CUE-ONLY). Numbers of significant neurons were subjected to a Bonferroni-corrected ($n = 4$) binomial test; asterisks indicate significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

Publication 2: Neuronal Codes for Arithmetic Rule Processing in the Human Brain

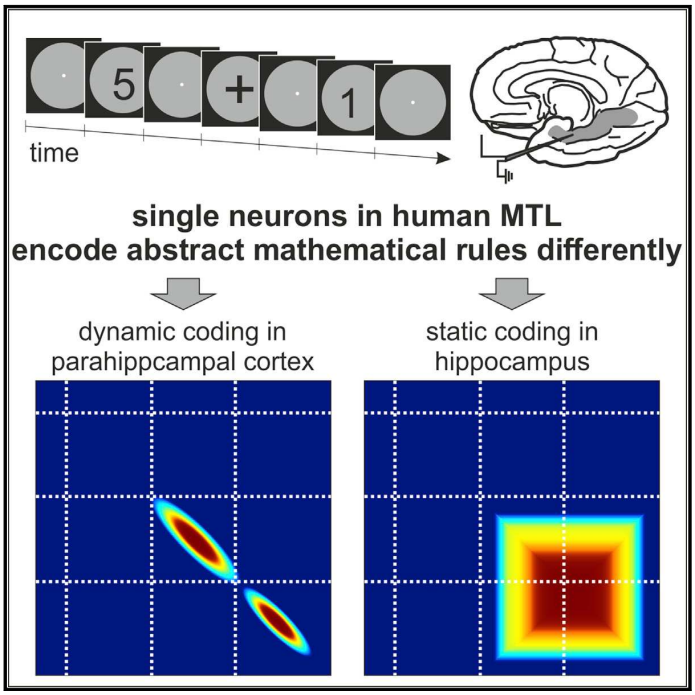
Kutter E.F., Bostroem J., Elger C.E., Nieder A., & Mormann F. (2022) Neuronal Codes for Arithmetic Rule Processing in the Human Brain. *Current Biology* **32(6)**: 1275–1284. DOI: 10.1016/j.cub.2022.01.054

Article

Current Biology

Neuronal codes for arithmetic rule processing in the human brain

Graphical abstract



Highlights

- Single neurons in the human MTL show abstract codes for addition and subtraction
- Time-resolved decoding analysis shows a dynamic code in the parahippocampal cortex
- The hippocampus shows a static code based on persistently rule-selective neurons
- Different codes suggest different cognitive functions of MTL regions in arithmetic

Authors

Esther F. Kutter, Jan Boström, Christian E. Elger, Andreas Nieder, Florian Mormann

Correspondence

andreas.nieder@uni-tuebingen.de (A.N.), florian.mormann@ukbonn.de (F.M.)

In brief

Kutter et al. demonstrate abstract and notation-independent codes for addition and subtraction in neuronal populations in the human medial temporal lobe (MTL). A dynamic code in the parahippocampal cortex contrasts with a static code in the hippocampus, suggesting different cognitive functions for these MTL regions in arithmetic.

Kutter et al., 2022, Current Biology *32*, 1275–1284
March 28, 2022
https://doi.org/10.1016/j.cub.2022.01.054

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

Neuronal codes for arithmetic rule processing in the human brain

Esther F. Kutter,[1,2] Jan Boström,[3] Christian E. Elger,[1] Andreas Nieder,[2,4,*] and Florian Mormann[1,4,5,6,*]
[1]Department of Epileptology, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany
[2]Animal Physiology, Institute of Neurobiology, University of Tübingen, 72076 Tübingen, Germany
[3]Department of Neurosurgery, University of Bonn Medical Center, Sigmund-Freud-Str. 25, 53105 Bonn, Germany
[4]These authors contributed equally
[5]Twitter: @humansingleunit
[6]Lead contact
*Correspondence: andreas.nieder@uni-tuebingen.de (A.N.), florian.mormann@ukbonn.de (F.M.)
https://doi.org/10.1016/j.cub.2022.01.054

SUMMARY

Arithmetic is a cornerstone of scientifically and technologically advanced human culture, but its neuronal mechanisms are poorly understood. Calculating with numbers requires temporary maintenance and manipulation of numerical information according to arithmetic rules. We explored the brain mechanisms involved in simple arithmetic operations by recording single-neuron activity from the medial temporal lobe of human subjects performing additions and subtractions. We found abstract and notation-independent codes for addition and subtraction in neuronal populations. The neuronal codes of arithmetic in different brain areas differed drastically. Decoders applied to time-resolved recordings demonstrate a static code in hippocampus based on persistently rule-selective neurons, in contrast to a dynamic code in parahippocampal cortex originating from neurons carrying rapidly changing rule information. The implementation of abstract arithmetic codes suggests different cognitive functions for medial temporal lobe regions in arithmetic.

INTRODUCTION

Mental arithmetic is an intricate skill and a hallmark of our scientifically advanced culture. Calculating with numbers requires semantic knowledge about numbers, online maintenance of numerical values, and their goal-directed transformation according to calculation rules. Therefore, mental arithmetic engages multiple brain systems, including those for the semantic representation of numeric values, the learning and memory of mathematical principles, and the cognitive control of mental operations.[1–4]

Studies in humans[1,2,5] and nonhuman primates[6–9] have indicated parts of the parietal and prefrontal cortices as core number representation and manipulation system. In particular, arithmetically selective brain areas have been identified in the parietal cortex of patients using intracranial electrocorticography (ECoG) recordings that measure summed and synchronized postsynaptic potentials (bulk tissue mass potentials).[10,11] Moreover, direct electrical stimulation studies in human patients have shown a specific arrest of counting and calculation performance during transient perturbation of parietal and frontal regions.[12–14] The latter investigations, in particular, suggest a causal involvement of parietal and frontal cortical regions in mental arithmetic.

However, recent findings implicate a wider cortical number network beyond parietal and frontal association cortices, also integrating the temporal lobe. Direct evidence resulted from ECoG studies in human patients; these recordings reconfirmed the presence of addition-selective locations not only in the posterior parietal cortex but also in the ventral temporal cortex.[10,11] In addition, functional neuroimaging implicated medial temporal lobe regions in the development of arithmetic fact knowledge,[3,15–17] including knowledge about arithmetic operators.[18] Moreover, performance enhancements in arithmetic fact retrieval are related to functional connectivity in hippocampal-neocortical circuits, including hippocampal-frontal[16,19] and hippocampal-parietal[16] connectivity. Hippocampal volume and functional connectivity of the hippocampus with dorsolateral and ventrolateral prefrontal cortices predict math tutoring success in children,[20] and reduced parahippocampal gray matter is associated with math learning disabilities (“developmental dyscalculia”).[21] Finally, we have recently shown directly by intracranial single-neuron recordings that the human medial temporal lobe (MTL) contains neurons that selectively respond to numerical values of different (symbolic and nonsymbolic) visual formats.[22] Here, we explored how single neurons in the human MTL represent the arithmetic addition and subtraction rules presented in different symbolic notations.

As a neuronal representation of the abstract rules applied to perceptual categories, rule-selective neurons have been identified in nonhuman primates.[23] They increase firing rates when a subject follows a specific rule but remain silent for alternative rules.[24,25] To bridge longer working memory delays necessary for mental arithmetic, two fundamentally different neuronal codes are conceivable: neurons might be tuned to a specific calculation rule and maintain this representation over long time periods via persistent firing. In this case, a decoder (a statistical classifier) trained on neuronal activity during a brief moment can successfully generalize across different time points. This

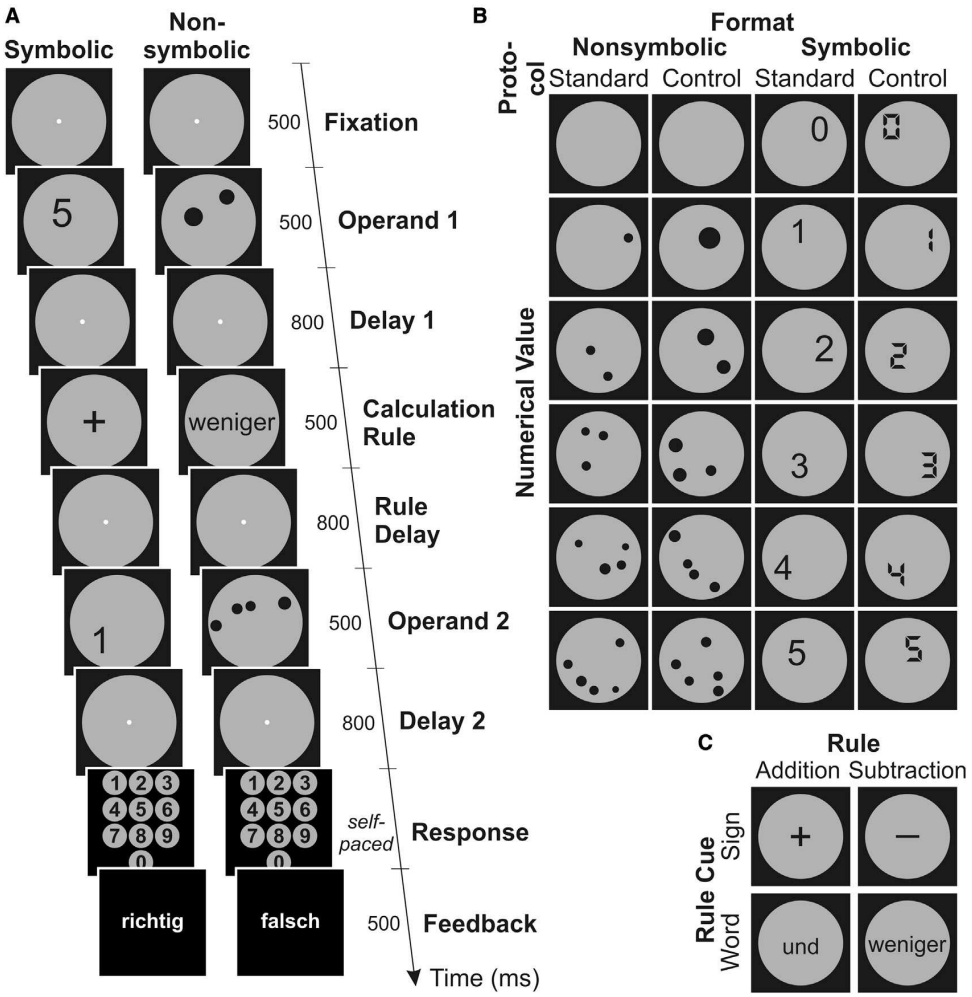


Figure 1. Behavioral task and example stimuli

(A) Experimental design of the calculation task. After visual fixation on the screen, stimuli were presented sequentially in the order operand 1—operator—operand 2. Each stimulus phase was followed by a brief delay. Afterward, the subjects were required to indicate the result of the calculation (ranging from 0 to 9) on an Arabic numeral panel and subsequently received feedback indicating whether the result was correct ("richtig") or false ("falsch").

(B) Example number stimuli for the nonsymbolic (numerosity) and symbolic format (numeral) for standard and control protocols. Numerical values of operand 1 ranged from 1 to 5; those of operand 2 ranged from 0 to 5.

(C) Example stimuli for the different mathematical rules, indicated by arithmetic signs ("+," "–") and written words ("und" [add], "weniger" [subtract]).

type of coding is known as static coding. Alternatively, neurons may fire sparsely and rapidly change tuning to calculation rules over time.[26] Under this scenario, a decoder trained on neuronal activity during one time point cannot generalize to the next. This has become known as dynamic coding.[27] By applying decoders to time-resolved recordings, we probe the codes for abstract arithmetic rules in the human MTL.

RESULTS

We asked nine human participants to perform simple addition and subtraction tasks on a computer display with operand values ranging from 0 to 5 (Figure 1A, see STAR Methods). Both operands were displayed with equal probability either as dot numerosities (nonsymbolic) or Arabic numerals (symbolic). Numerosities were shown in standard (variable dot size and arrangement) and control (constant total dot area and dot density) displays to control for non-numerical visual parameters; Arabic numerals were shown in two (standard and control) font types to ensure the generalization of symbols across visual shapes (Figure 1B). Addition and subtraction rules were instructed by two different notations, either as arithmetic signs (+, –) or written words (German "und" and "weniger," indicating "add" and "subtract"). The two rule notations allowed us to later

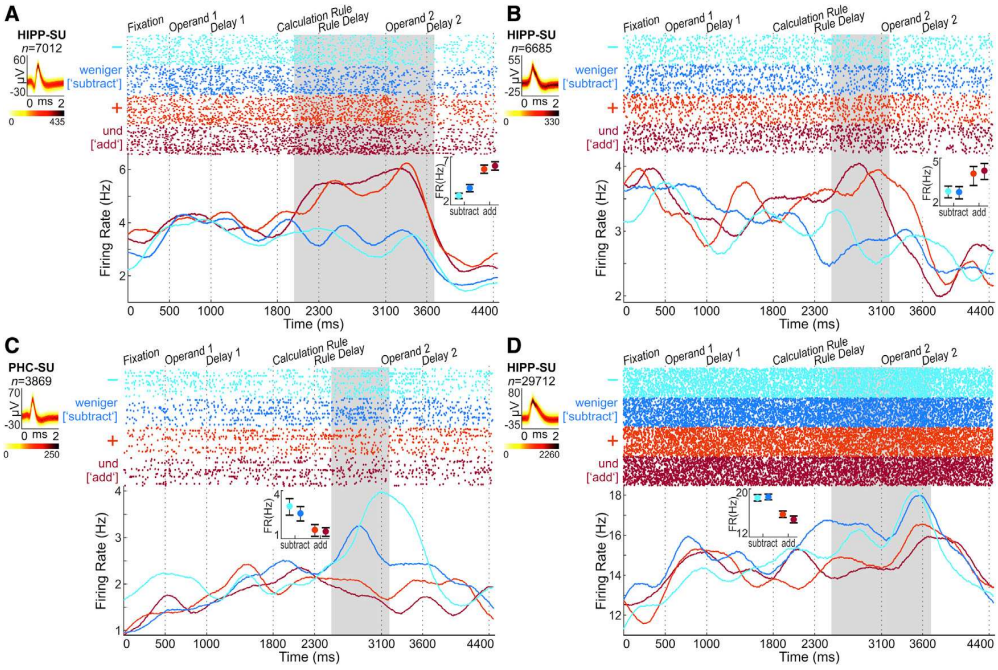


Figure 2. Neural responses of rule-selective neurons

(A–D) Across-trial responses of four example neurons responding with increased firing rate to the "addition" rule (A, B) and the "subtraction" rule (C, D), regardless of the concrete cue indicating the rule. The left small panels depict a density plot of the recorded action potentials (color darkness indicates the number of overlapping wave forms according to color scale at the bottom). Top panels show single-cell dot-raster plots for many repetitions of the rule cue (each line represents a trial and each dot represents an action potential, color coded according to the two rules and the two rule cues), and averaged instantaneous firing rates (spike-density histograms) are shown below. Blueish colors depict subtraction (for two different rule cues); reddish colors correspond to addition. Insets show average activity per rule cue during the rule discrimination period (gray shaded area), as defined by statistical significance in the ANOVA. Error bars denote SEM.

dissociate neural activity related to the visual properties of the cue (sign or word) from the abstract rule that it represented (addition versus subtraction) (Figure 1C). The participants' average performance was close to perfect (98.5% ± 0.6%) and comparable between addition and subtraction ($p = 0.97$; t test).

Single neurons respond to calculation rules

We recorded the action potentials of a total of 585 single neurons in the MTLs of the participants performing the calculation tasks: 126 neurons in parahippocampal cortex (PHC), 199 neurons in hippocampus (HIPP), 107 neurons in entorhinal cortex (EC), and 153 neurons in amygdala (AMY). As an obligatory but not sufficient prerequisite for mental calculation, MTL neurons were previously shown to represent numerical cardinality of the first operand.[22] We predicted that single neurons and neuronal populations also encode mentally performed additions and subtractions.

Using a multi-factor analysis of variance (ANOVA) (see STAR Methods), we first identified rule-selective neurons that selectively increased firing rates to either the addition or subtraction rule after the instruction of the calculation ("calculation rule") (Figure 1A). After the presentation of the calculation rule, addition-selective neurons enhanced firing whenever an addition was instructed (Figures 2A and 2B; reddish colors); whereas subtraction-selective neurons showed a specific increase in activity whenever a subtraction was cued (Figures 2C and 2D; blueish colors).

The proportion of neurons selectively tuned to calculation rules differed for different task periods and MTL areas (Figure 3). In the "calculation rule" period, a small but significant proportion of MTL neurons (4.8%; 28/585) was modulated by the arithmetic rule ($p < 0.001$; binomial test with $p_{chance} = 0.01$). Most of these neurons (3.5%; 20/585) showed activity that varied exclusively with the arithmetic rule ($p < 0.001$; binomial test), irrespective of the cue indicating that rule (i.e., no significant main effect for the factor cue, or any other factor) (Figure 3A, first column). Only PHC (7%; 9/126) and HIPP (4%; 7/199), but not EC and AMY, showed proportions of such exclusively rule-selective neurons higher than expected by chance ($p < 0.05$; binomial test with $p_{chance} = 0.01$, Bonferroni-corrected for multiple tests across areas, $n = 4$; Figures 3B and 3D). The overall proportion of rule-selective neurons increased in the "rule delay" period in which the value of operand 1 and the calculation instruction needed to be held in working memory to solve the task. Here, 6.0% (35/585) of MTL neurons were modulated by the arithmetic rule, with 5.3% (31/585) being exclusively rule-selective (both

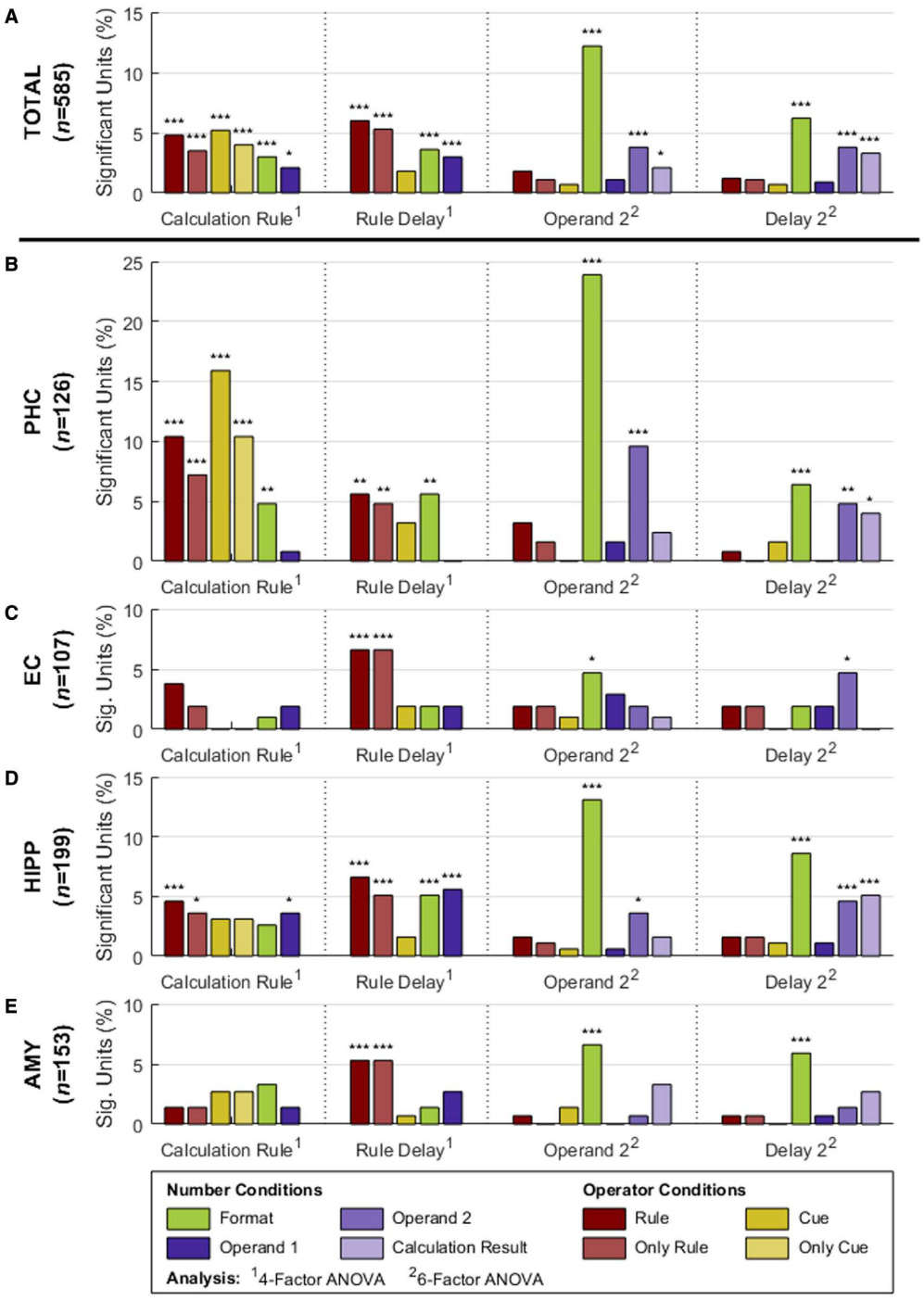


Figure 3. Neuronal selectivity of MTL single units

(A–D) Proportions of single units significant to different task factors for different MTL regions: (A) total population, (B) parahippocampal cortex, (C) entorhinal cortex, (D) hippocampus, and (E) amygdala. ANOVAs for the different task phases were evaluated at $\alpha = 0.01$. Neurons with an effect for "arithmetic rule," but no concurrent other main effects are termed "exclusively rule-selective" ("Only Rule"); same applies for factor "rule cue." Numbers of significant neurons were subjected to a Bonferroni-corrected (n = 4) binomial test; asterisks indicate significance (*p < 0.05, **p < 0.01, ***p < 0.001). See also Figures S1 and S2.

$p < 0.001$, binomial test; Figure 3A, second column). During this period, all MTL areas contained a significant proportion of exclusively calculation-rule-selective neurons (EC: 7%, 7/107; HIPP: 5%, 10/199; PHC: 5%, 6/126; and AMY: 5%, 8/153; all $p < 0.01$, Bonferroni-corrected binomial test) in addition to neurons coding other task-relevant factors (Figures 3B–3E). The proportion of rule-selective neurons dropped to chance level in the subsequent task periods (“operand 2” and “delay 2”) (Figure 3A). Instead, the neurons represented the task factors additionally introduced with trial progression (for instance, the numerical value of operand 2) (Figure 3, third and fourth columns). In sum, a small but significant proportion of neurons encoded arithmetic rules after the presentation and memorization of the calculation rule prior to the presentation of operand 2. A separate analysis of the neurons in the left and right hemisphere qualitatively confirmed these results for each hemisphere in all regions (Figure S1). Note that the proportions of neurons showing significant selectivity during different trial periods do not generally represent the same neuronal populations (see Figure S2).

Cue-independent representation of addition and subtraction rules

MTL neurons showed variation in the time point and duration of rule selectivity across the task period. With increasing task complexity later in the trial, they also exhibited selectivity to several task factors, seemingly multiplexing the different information required to solve the task. Therefore, we focused our attention on the collective properties of groups of neurons, or “neuronal populations.” This allowed us to read out (or “decode”) information not only from an individual neuron but from a population of neurons that the subjects can base their decisions on. In combination with decoding methods, such as statistical classifiers, this allows to predict the accuracy and abstractness of arithmetic rule representations on a trial-by-trial basis.

To explore whether this variable activity yielded a reliable read-out of arithmetic rules, we adopted a machine-learning approach. We trained support vector machine (SVM) classifiers to discriminate between addition trials (“und” [add] and “+” cues combined) and subtraction trials (“weniger” [subtract] and “–” cues combined) across trial periods based on firing rates (see STAR Methods). The classifiers were then tested on novel data from the same neurons to explore how well they could predict the rules based on the information extracted from trials used during training. Cluster permutation tests ($p < 0.05$) were used to identify the trial intervals of classification accuracies significantly above chance level (50%).

We found long time intervals for which rule information could be successfully decoded in all MTL regions (black horizontal bars in Figures 4A–4D). Consistent with the single-cell analysis, we observed strong and long-lasting effects in PHC (Figure 4A; two selective periods interrupted by a non-selective period) and HIPP (Figure 4C). In HIPP particularly, the calculation rule was continuously decoded with high accuracy from rule onset until the end of the trial (Figure 4C). By contrast, rather weak and short-lived significant classification performance was observed in EC (Figure 4B) and AMY (Figure 4D). Separate analysis of left and right hemisphere qualitatively confirmed these results for PHC and HIPP (Figure S3).

We wanted to find out whether the calculation rules could be decoded irrespective of the rule cues, as would be expected for abstract rule coding. Therefore, we explored classification accuracies for individual rule cues during the previously found significant time intervals shown in Figures 4A–4D. We trained an SVM classifier using the firing rates combined for both rule cues per calculation rule in the individual significant windows for each MTL area. We then tested whether the SVM could predict the correct calculation rule from novel trials based on either one of the two applied cues per calculation rule (i.e., signs and words).

We found that classifier performance for addition and subtraction across word and sign rule cues was significant in all MTL areas ($p < 0.05$; permutation test compared with shuffled data labels). Highest classification accuracies were found in HIPP (addition: 74% and subtraction: 73%; Figure 4G) and PHC (addition: 65% and subtraction: 68%; averaged across both significant time windows; Figure 4E) followed by EC (addition: 65% and subtraction: 61%; Figure 4F) and AMY (addition: 58% and subtraction: 63%; Figure 4H). In AMY, performance was mainly due to accurate encoding of one specific cue (the “weniger” [subtract] cue), whereas classification accuracies were relatively low for the other three cues. Overall, information about the calculation rules was encoded irrespective of the rule cues prompting addition and subtraction, respectively.

Cross-temporal calculation rule decoding (static-dynamic)

Next, we explored the neuronal codes of arithmetic rules. By applying decoders to time-resolved recordings, we asked whether the code remained stable across trial time or rather changed dynamically (Figures 4I–4L). To this end, we performed a temporal cross-training analysis. We trained SVM classifiers on the firing rates from a given time point and tested them during other time points of the trial. This analysis was again performed separately for the four MTL areas, and the accuracy results were plotted in a confusion matrix spanning the trial times of classifier training against the trial times of classifier testing.

In PHC, we observed that classifiers trained during a specific time interval after rule cue presentation were only able to decode the arithmetic rule in the same time interval (Figure 4I). This resulted in high classification accuracies only along the main diagonal of the confusion matrix. The classifiers’ inability to generalize the calculation rules across trial time periods indicates a dynamic neuronal code in PHC based on neurons that rapidly change their tuning properties with time.

A rather different picture emerged for HIPP (Figure 4K). Significant cross-temporal generalization from the end of the rule cue period all the way up to the end of the trial was present. A classifier trained on firing rates observed, e.g., during the rule delay, was still able to decode the calculation rule when tested on activity recorded during the operand 2 or even delay 2 phases. This resulted in a square-like accuracy pattern in the cross-temporal confusion matrix. This pattern argues for a static neuronal code in HIPP based on tuned neurons that remained stable across time throughout the trial.

In EC, cross-temporal generalization was weak and observed only to a small extent (Figure 4J). A square-like accuracy pattern emerged only around the rule delay. This suggests relatively

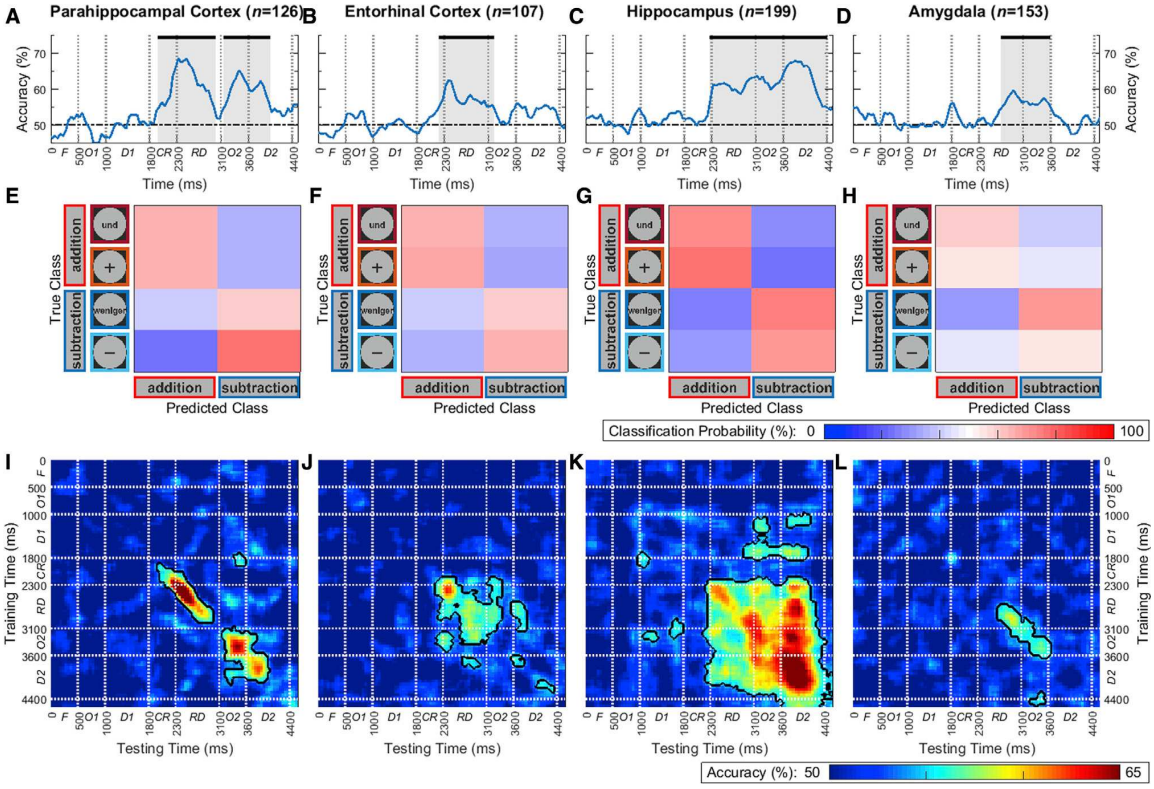


Figure 4. Rule decoding using a support vector machine (SVM) classifier. Decoding performance for the four different MTL regions (columns)

(A–D) Classification accuracy for decoding arithmetic rule information when training an SVM classifier on the instantaneous firing rates across the trial period. The dashed line represents chance level (50% for two classes). Black bars above the data and areas shaded in gray indicate significance ($p < 0.05$) when testing against performance for SVMs trained on shuffled data in a permutation test. Abbreviations at the axes indicate task phases: *F*, fixation; *O1*, operand 1; *D1*, delay 1; *CR*, calculation rule; *RD*, rule delay; *O2*, operand 2; and *D2*, delay 2.

(E–H) Confusion matrices derived from training an SVM classifier on firing rates averaged across the significant time windows in (A–D), respectively. (E) shows the average of the confusion matrices obtained for each significant window (depicted in A).

(I–L) Accuracy when training an SVM classifier at a given time point of the trial and testing on another one (the main diagonal of the matrix corresponds to the curve in [A–D]). Black contours indicate significance ($p < 0.05$) in a permutation test. See also Figures S3–S5.

stable calculation rule coding for as long as only the calculation rule was kept in working memory and before the second operand was presented. In AMY, neurons did not encode arithmetic rules abstractly (see Figure 4H). In addition, the cross-temporal classifier analysis showed only a mild accuracy diagonal during the rule delay and the operand 2 periods (Figure 4L). Both findings preclude statements about coding dynamics in AMY.

The observed response patterns, in particular a dynamic code in PHC contrasted by a static code in HIPP, were still present after equalizing the numbers of neurons for each MTL area (Figure S4; see STAR Methods). Information about the rule cues were only encoded during cue presentation in PHC (Figure S5), which further confirmed the abstractness of the calculation rule coding.

Cross-notation decoding of addition and subtraction

The previous analyses showed that the population of neurons differentiated between addition and subtraction rules and indicated that calculation rules were encoded irrespective of the rule cues. As a final step, we put this observation to the test and explored whether MTL neurons generalize calculation rules across rule notations.

We performed a time-resolved sliding-window decoding analysis and trained a SVM classifier on the trials of one rule notation and tested it on the other rule notation for the same calculation rule. First, we used all the word trials (“und” [add] and “weniger” [subtract]) as training data and all the sign trials [“+” and “–”] as test data [henceforth called “word → sign”]) and performed the same analyses as before (i.e., temporal cross-training classification and verification via fixed-window analysis). Then, we analyzed the generalization in the opposite direction, i.e., using all sign trials as training data and all word trials as test data (in the following called “sign → word”) and repeating the whole procedure. Generalization was judged successful if (1) synchronous intervals of significant classification for both directions of generalization were found (Figures 5A–5D)

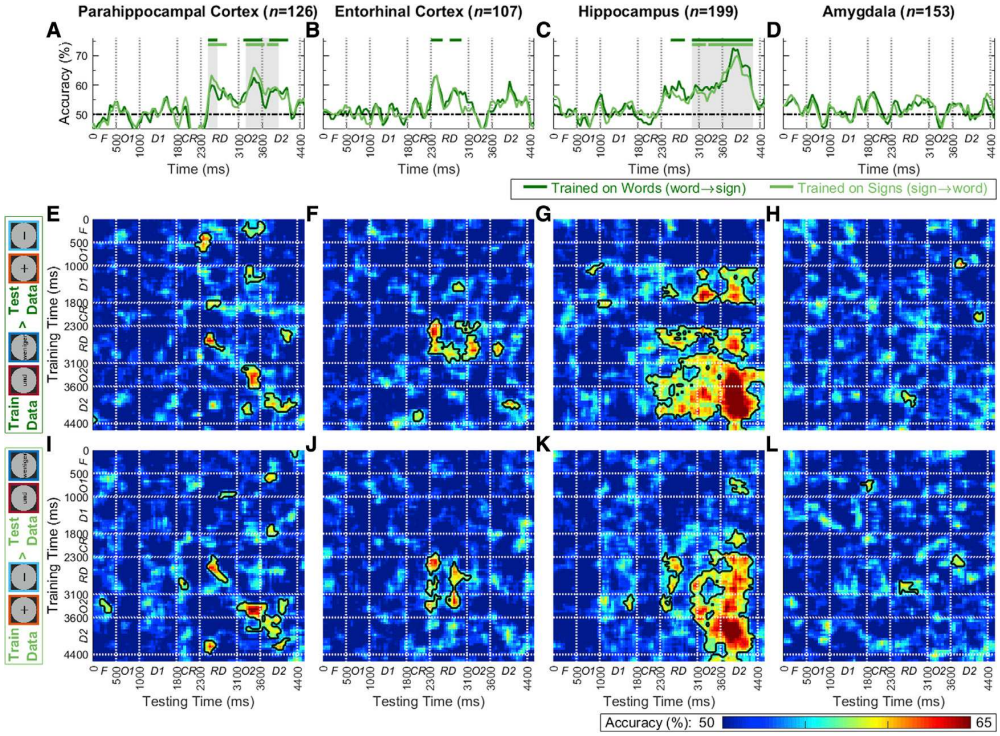


Figure 5. Generalization between arithmetic rule notations using an SVM classifier. Classifier performance for the four different MTL regions (columns)

(A–D) Classification accuracy when training an SVM on the instantaneous firing rates across the trial period for both directions of generalization. The dashed line represents chance level (50% for two classes). Light and dark green bars above the data indicate significance ($p < 0.05$) in a permutation test for both test directions ("word → sign" and "sign → word," respectively). The areas shaded in gray indicate the synchronous time windows used for the fixed-window classification analysis. Abbreviations at the axes indicate task phases: *F*, fixation; *O1*, operand 1; *D1*, delay 1; *CR*, calculation rule; *RD*, rule delay; *O2*, operand 2; and *D2*, delay 2.

(E–H) Accuracy in temporal cross-training analysis when a classifier is trained on trials showing word rule cues and tested on trials showing sign rule cues (the main diagonals of the matrices correspond to the dark green curves in [A–D]). Black contours indicate significance ($p < 0.05$) when testing against performance for SVMs trained on shuffled data in a permutation test.

(I–L) Accuracy in temporal cross-training analysis when a classifier is trained on trials showing sign cues and tested on trials showing word cues (the main diagonals of the matrices correspond to the light green curves in [A–D]). Same conventions as in (E–H).

and (2) performance in these synchronous intervals was significant (permutation test, $p < 0.05$) for each arithmetic rule after averaging classification accuracies across both directions of generalization.

Significant cross-notation rule decoding was present in PHC (Figure 5A) and HIPP (Figure 5C). In both areas, extended and overlapping intervals of significant rule classifications for both test directions ("word → sign" and "sign → word") emerged after rule cue offset and up to the end of the trial (significant phases are indicated by light and dark green horizontal bars in Figures 5A and 5C, with synchronous intervals indicated by gray shaded areas). In PHC, the accuracy of transfer was 57% for addition and 64% for subtraction (Figure 5A). In HIPP, transfer was even better and reached an average accuracy of 73% for addition and 69% for subtraction (Figure 5C). In both PHC and HIPP, the transfer for both calculation rules and both test directions were individually significant and thus pooled.

By contrast, the cross-notation decoding of calculation rules failed in AMY and EC. In AMY, classification accuracy remained at chance level throughout the whole trial for both directions of generalization (Figure 5D). In EC, cross-notation decoding briefly transferred for the direction "word → sign," but not for the direction "sign → word," and therefore failed our generalization criterion (Figure 5B). Thus, activity in PHC and HIPP did generalize arithmetic rule information across notations, whereas AMY and EC did not.

To explore the dynamics of rule codes during cross-notation generalization, we again employed a temporal cross-training analysis separately for the two generalization directions. The resulting confusion matrices confirmed the earlier findings (albeit with weaker effect size due to the reduced data dimensionality). In HIPP, they showed a static rule code for both notation generalization directions from rule cue offset to the end of the trial (indicated by the square-like significant classification pattern) (Figures 5G and 5K). By contrast, a dynamic code

emerged for both notation generalization directions in PHC, as evidenced by significant classification performance along the main diagonal of the matrices (Figures 5E and 5I). The findings for EC and AMY were unreliable due to the lack of rule-notation generalization found in these areas (Figures 5B and 5D). In summary, neuron populations in both PHC and HIPP generalized between calculation rule notation but exhibited fundamentally different rule codes.

DISCUSSION

Our findings demonstrate that the activity of neurons in the MTL carries sufficient information to allow a statistical classifier to discriminate between addition and subtraction instructions during mental calculation. After having been trained on activity during mental processing of one rule notation, decoding generalized to another notation cueing the same arithmetic operation. This generalization observed for arithmetic signs and words implies an abstract and notation-independent representation of arithmetic rules. Therefore, our research unveils a neuronal correlate for mental arithmetic, which generalizes between calculation tasks involving learned mathematical symbols.

Number and calculation recruit the MTL

Our discovery of arithmetic rule selectivity in MTL is in agreement with a growing body of studies that suggests that number and calculation recruit an interconnected network of cortical areas, including parts of the temporal lobe. We have recently shown directly by intracranial single-neuron recordings that the human MTL contains a significant percentage of neurons selectively tuned to numerical values.[22] This finding in humans concurs with the numerical tuning of hippocampal neurons in nonhuman primates.[28] Moreover, using functional imaging, the MTL has been shown to be involved in arithmetic skill acquisition and memory-based problem-solving strategies during childhood.[15,16,20] Similarly, ECoG studies in human patients reported addition-selective locations not only in the posterior parietal cortex but also in the ventral temporal cortex.[10,11] Together, these data implicate a wider cortical number network beyond parietal and frontal association cortices, also integrating the temporal lobe.

The MTL is suited to transform and manipulate representations of incoming numerical information.[22] It is highly interconnected with the frontal and parietal areas[29,30] that constitute the core number system involved in perceiving the number of sensory stimuli.[2,7,31] The prefrontal lobe, in particular, is associated with representing abstract rules and concepts, information that can be directly accessed by MTL.[23] Therefore, MTL could mediate the transformation of perceived numerical information in a working memory buffer. Interestingly, as an associative brain area, the MTL also contains sensorimotor neurons that are activated by hand-grasping observation and execution.[32] This opens the possibility that MTL may play a role in the sensorimotor translation of perceived and produced number,[33] a speculation yet to be explored.

Numerical rule-selective neurons

Although the observed rule-selective neurons encode arithmetic rules, this is not to say that some of these neurons may not also become engaged in encoding other types of rules that we have not explored. However, recordings show that rule-selective neurons in the nonhuman primate exhibit a substantial degree of specialization and preferentially respond only to quantitative rules applied to a specific magnitude type.[34] Therefore, we think that the majority of rule-selective neurons specifically and genuinely encode arithmetic rules.

So far, the neuronal correlates of addition and subtraction have not been studied in monkeys. However, what has been investigated is how single neurons respond to more basic mathematical operations, namely, "greater than" (>) and "less than" (<) operations.[6,24,34] In each trial, monkeys had to flexibly switch between these two rules according to rule cues and had to choose either a larger (in the case of the "greater than" rule) or smaller numerosity (in the case of the "smaller than" rule) than the one they had seen in the beginning of a trial. Recording from frontal and parietal areas, we found single neurons that responded selectively (by increased firing rates) to one of the two rules. Rule selectivity was stronger and more abundant in the frontal lobe than in the parietal lobe.[24] In monkeys, the frontal association areas are thus more important when it comes to nonsymbolic mathematical rules. This is consistent with imaging results in humans, where areas specific to calculation rather than simple number comparison are primarily found in the frontal lobe.[2]

The numerical coding capacities of such neurons do not seem to be fixed. Although neurons selective to numerical rules were recorded in brain areas in which previous studies had shown a relatively high proportion of numerosity-selective neurons,[7–9] the number of neurons representing pure numerical values were reduced when monkeys applied numerical rules.[24,34] This indicates that the neurons in the fronto-parietal number network dynamically encode different types of numerical information as a function of task demands. In the same vein, flexible numerical coding may apply to the MTL during a (top-down) interplay with the frontal lobe, depending on the precise mathematical task at hand.

Static and dynamic calculation codes

Mental calculation is a classic working memory task, and although working memory has traditionally been attributed to the prefrontal cortex,[35] more recent data suggest that the MTL may also be important in working memory tasks[36–38] and that it is part of a brain-wide network subserving working memory.[39]

Previous intracranial recording studies show that the delay activity of a selection of MTL neurons correlated with memory load and predict the successful retrieval of working memory contents.[39,40] These neurons' persistent activation maintained the same stimulus preference throughout several seconds of temporal gaps. This type of activity with robust across-time generalization is characteristic of a static code that we also observed in the hippocampus during mental calculation.

In contrast to the static code in hippocampus, we observed a clear dynamic code in PHC when processing calculation rules. Such a dynamic code based on sparsely bursting neurons supports the theory of activity-silent working memory.[41] It proposes that working memory can also be supported by short-term changes in synaptic weights. Synaptic weight changes are involved in episodic memory, which is why activity-silent working memory might be reminiscent of—or

even part of—episodic memory.[42] This fits with the finding that areas of the MTL are not only critically implicated in episodic memory but also important during working memory tasks.[36–40,43] Direct observation of neuronal reactivation after complete activity silence has recently been reported in a different working memory task in the human MTL.[44] Static and dynamic codes are not incompatible. Stable persistent activation with robust across-time generalization can exist in the presence of dynamically changing neuronal representations.[45,46]

Neuron recordings in human[39,40,43] and nonhuman primates,[47–49] as well as computational modeling,[50–52] suggest different cognitive functions for these two codes for working memory: although a dynamic code seems to suffice for short maintenance of more implicit information in memory, the intense mental manipulation of the attended working memory contents may require a static code. Following this logic, parahippocampal cortex may represent a short-term memory of the arithmetic rule, whereas downstream hippocampus may ''do the math'' and process numbers according to the arithmetic rule at hand. More fine-grained analyses, ideally combined with perturbation approaches,[53] will help to decipher the individual roles of brain areas and neuronal codes in mental arithmetic.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Experimental Task and Stimuli
  - Neurophysiological Recording
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Neuronal Analysis of Variance (ANOVA)
  - Support Vector Machine (SVM) Classification
  - Generalization of SVM Classification across Rule Cue Notations

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2022.01.054.

ACKNOWLEDGMENTS

We thank all patients for their participation. This research was supported by the Volkswagen Foundation and the German Research Council (SPP 2205, SFB 1089, Mo 930/4-2, Ni 618/10-1, Ni 618/11-1).

AUTHOR CONTRIBUTIONS

A.N. and F.M. designed the study. C.E.E. and F.M. recruited patients. J.B. and F.M. implanted the electrodes. E.F.K. and F.M. collected the data. E.F.K. analyzed the data with contributions from A.N. and F.M.. A.N., E.F.K., and F.M. wrote the paper. All authors discussed the results and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 4, 2020
Revised: December 20, 2021
Accepted: January 20, 2022
Published: February 14, 2022

REFERENCES

1. Amalric, M., and Dehaene, S. (2016). Origins of the brain networks for advanced mathematics in expert mathematicians. Proc. Natl. Acad. Sci. USA *113*, 4909–4917.
2. Arsalidou, M., and Taylor, M.J. (2011). Is 2+2=4? Meta-analyses of brain areas needed for numbers and calculations. NeuroImage *54*, 2382–2393.
3. Menon, V. (2016). Memory and cognitive control circuits in mathematical cognition and learning. Prog. Brain Res. *227*, 159–186.
4. Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. Annu. Rev. Neurosci. *32*, 185–208.
5. Piazza, M., Pinel, P., Le Bihan, D., and Dehaene, S. (2007). A magnitude code common to numerosities and number symbols in human intraparietal cortex. Neuron *53*, 293–305.
6. Bongard, S., and Nieder, A. (2010). Basic mathematical rules are encoded by primate prefrontal cortex neurons. Proc. Natl. Acad. Sci. USA *107*, 2277–2282.
7. Nieder, A. (2016). The neuronal code for number. Nat. Rev. Neurosci. *17*, 366–382.
8. Ramirez-Cardenas, A., Moskaleva, M., and Nieder, A. (2016). Neuronal representation of numerosity zero in the primate parieto-frontal number network. Curr. Biol. *26*, 1285–1294.
9. Viswanathan, P., and Nieder, A. (2015). Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. Curr. Biol. *25*, 1259–1269.
10. Daitch, A.L., Foster, B.L., Schrouff, J., Rangarajan, V., Kaşikçi, I., Gattas, S., and Parvizi, J. (2016). Mapping human temporal and parietal neuronal population activity and functional coupling during mathematical cognition. Proc. Natl. Acad. Sci. USA *113*, E7277–E7286.
11. Pinheiro-Chagas, P., Daitch, A., Parvizi, J., and Dehaene, S. (2018). Brain mechanisms of arithmetic: A crucial role for ventral temporal cortex. J. Cogn. Neurosci. *30*, 1757–1772.
12. Della Puppa, A., De Pellegrin, S., d'Avella, E., Gioffrè, G., Munari, M., Saladini, M., Salillas, E., Scienza, R., and Semenza, C. (2013). Right parietal cortex and calculation processing: intraoperative functional mapping of multiplication and addition in patients affected by a brain tumor. J. Neurosurg. *119*, 1107–1111.
13. Roux, F.-E., Boukhatem, L., Draper, L., Sacko, O., and Démonet, J.-F. (2009). Cortical calculation localization using electrostimulation. J. Neurosurg. *110*, 1291–1299.
14. Semenza, C., Salillas, E., De Pallegrin, S., and Puppa, A.D. (2017). Balancing the 2 hemispheres in simple calculation: evidence from direct cortical electrostimulation. Cereb. Cortex N. Y. N 1991 *27*, 4806–4814.
15. De Smedt, B., Holloway, I.D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetical fluency. NeuroImage *57*, 771–781.
16. Qin, S., Cho, S., Chen, T., Rosenberg-Lee, M., Geary, D.C., and Menon, V. (2014). Hippocampal-neocortical functional reorganization underlies children's cognitive development. Nat. Neurosci. *17*, 1263–1269.
17. Peters, L., and De Smedt, B. (2018). Arithmetic in the developing brain: a review of brain imaging studies. Dev. Cogn. Neurosci. *30*, 265–279.
18. Mathieu, R., Epinat-Duclos, J., Léone, J., Fayol, M., Thevenot, C., and Prado, J. (2018). Hippocampal spatial mechanisms relate to the development of arithmetic symbol processing in children. Dev. Cogn. Neurosci. *30*, 324–332.

19. Cho, S., Metcalfe, A.W., Young, C.B., Ryali, S., Geary, D.C., and Menon, V. (2012). Hippocampal-prefrontal engagement and dynamic causal interactions in the maturation of children's fact retrieval. J. Cogn. Neurosci. *24*, 1849–1866.
20. Supekar, K., Swigart, A.G., Tenison, C., Jolles, D.D., Rosenberg-Lee, M., Fuchs, L., and Menon, V. (2013). Neural predictors of individual differences in response to math tutoring in primary-grade school children. Proc. Natl. Acad. Sci. USA *110*, 8230–8235.
21. Rykhlevskaia, E., Uddin, L.Q., Kondos, L., and Menon, V. (2009). Neuroanatomical correlates of developmental dyscalculia: combined evidence from morphometry and tractography. Front. Hum. Neurosci. *3*, 51.
22. Kutter, E.F., Bostroem, J., Elger, C.E., Mormann, F., and Nieder, A. (2018). Single neurons in the human brain encode numbers. Neuron *100*, 753–761.e4.
23. Mansouri, F.A., Freedman, D.J., and Buckley, M.J. (2020). Emergence of abstract rules in the primate brain. Nat. Rev. Neurosci. *21*, 595–610.
24. Vallentin, D., Bongard, S., and Nieder, A. (2012). Numerical rule coding in the prefrontal, premotor, and posterior parietal cortices of macaques. J. Neurosci. *32*, 6621–6630.
25. Wallis, J.D., Anderson, K.C., and Miller, E.K. (2001). Single neurons in prefrontal cortex encode abstract rules. Nature *411*, 953–956.
26. Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. Neuron *78*, 364–375.
27. King, J.-R., and Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. *18*, 203–210.
28. Opris, I., Santos, L.M., Gerhardt, G.A., Song, D., Berger, T.W., Hampson, R.E., and Deadwyler, S.A. (2015). Distributed encoding of spatial and object categories in primate hippocampal microcircuits. Front. Neurosci. *9*, 317.
29. Goldman-Rakic, P.S., Selemon, L.D., and Schwartz, M.L. (1984). Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. Neuroscience *12*, 719–743.
30. Suzuki, W.A. (2009). Comparative analysis of the cortical afferents, intrinsic projections, and interconnections of the parahippocampal region in monkeys and rats. In The Cognitive Neurosciences, Fourth Edition (Massachusetts Institute of Technology), pp. 659–674.
31. Nieder, A. (2004). The number domain- can we count on parietal cortex? Neuron *44*, 407–409.
32. Mukamel, R., Ekstrom, A.D., Kaplan, J., Iacoboni, M., and Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. Curr. Biol. *20*, 750–756.
33. Anobile, G., Arrighi, R., Castaldi, E., and Burr, D.C. (2021). A sensorimotor numerosity system. Trends Cogn. Sci. *25*, 24–36.
34. Eiselt, A.-K., and Nieder, A. (2013). Representation of abstract quantitative rules applied to spatial and numerical magnitudes in primate prefrontal cortex. J. Neurosci. *33*, 7526–7534.
35. Goldman-Rakic, P.S. (1995). Cellular basis of working memory. Neuron *14*, 477–485.
36. Goodrich, R.I., Baer, T.L., Quent, J.A., and Yonelinas, A.P. (2019). Visual working memory impairments for single items following medial temporal lobe damage. Neuropsychologia *134*, 107227.
37. Koen, J.D., Borders, A.A., Petzold, M.T., and Yonelinas, A.P. (2017). Visual short-term memory for high resolution associations is impaired in patients with medial temporal lobe damage. Hippocampus *27*, 184–193.
38. Olson, I.R., Moore, K.S., Stark, M., and Chatterjee, A. (2006). Visual working memory is impaired when the medial temporal lobe is damaged. J. Cogn. Neurosci. *18*, 1087–1097.
39. Kornblith, S., Quian Quiroga, R., Koch, C., Fried, I., and Mormann, F. (2017). Persistent single-neuron activity during working memory in the human medial temporal lobe. Curr. Biol. *27*, 1026–1032.
40. Kamiński, J., Sullivan, S., Chung, J.M., Ross, I.B., Mamelak, A.N., and Rutishauser, U. (2017). Persistently active neurons in human medial frontal and medial temporal lobe support working memory. Nat. Neurosci. *20*, 590–601.
41. Stokes, M.G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. Trends Cogn. Sci. *19*, 394–405.
42. Beukers, A.O., Buschman, T.J., Cohen, J.D., and Norman, K.A. (2021). Is activity silent working memory simply episodic memory? Trends Cogn. Sci. *25*, 284–293.
43. Boran, E., Fedele, T., Klaver, P., Hilfiker, P., Stieglitz, L., Grunwald, T., and Sarnthein, J. (2019). Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. Sci. Adv. *5*, eaav3687.
44. Bausch, M., Niediek, J., Reber, T.P., Mackay, S., Boström, J., Elger, C.E., and Mormann, F. (2021). Concept neurons in the human medial temporal lobe flexibly represent abstract relations between concepts. Nat. Commun. *12*, 6164.
45. Murray, J.D., Bernacchia, A., Roy, N.A., Constantinidis, C., Romo, R., and Wang, X.J. (2017). Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. Proc. Natl. Acad. Sci. USA *114*, 394–399.
46. Spaak, E., Watanabe, K., Funahashi, S., and Stokes, M.G. (2017). Stable and dynamic coding for working memory in primate prefrontal cortex. J. Neurosci. *37*, 6503–6516.
47. Buschman, T.J., Siegel, M., Roy, J.E., and Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. Proc. Natl. Acad. Sci. USA *108*, 11252–11255.
48. Mendoza-Halliday, D., Torres, S., and Martinez-Trujillo, J.C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. Nat. Neurosci. *17*, 1255–1262.
49. Sarma, A., Masse, N.Y., Wang, X.-J., and Freedman, D.J. (2016). Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. Nat. Neurosci. *19*, 143–149.
50. Masse, N.Y., Yang, G.R., Song, H.F., Wang, X.-J., and Freedman, D.J. (2019). Circuit mechanisms for the maintenance and manipulation of information in working memory. Nat. Neurosci. *22*, 1159–1167.
51. Mongillo, G., Barak, O., and Tsodyks, M. (2008). Synaptic theory of working memory. Science *319*, 1543–1546.
52. Wolff, M.J., Jochim, J., Akyürek, E.G., and Stokes, M.G. (2017). Dynamic hidden states underlying working-memory-guided behavior. Nat. Neurosci. *20*, 864–871.
53. Lee, D.K., Fedorenko, E., Simon, M.V., Curry, W.T., Nahed, B.V., Cahill, D.P., and Williams, Z.M. (2018). Neural encoding and production of functional morphemes in the posterior temporal lobe. Nat. Commun. *9*, 1877.
54. Niediek, J., Boström, J., Elger, C.E., and Mormann, F. (2016). Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours. PLoS ONE *11*, e0166598.
55. Mormann, F., Kornblith, S., Quiroga, R.Q., Kraskov, A., Cerf, M., Fried, I., and Koch, C. (2008). Latency and selectivity of single neurons indicate hierarchical processing in the human medial temporal lobe. J. Neurosci. *28*, 8865–8872.
56. Chang, C.-C., and Lin, C.-J. (2011). LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. *2*, 1–27.
57. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods *164*, 177–190.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and Algorithms		
Cheetah software	Neuralynx Inc.	https://neuralynx.com/software/cheetah
Combinato spike sorting software	Niediek et al.[54]	https://github.com/jniediek/combinato
MATLAB R2017a	MathWorks	https://de.mathworks.com/
Psychtoolbox	Brainard (1997)	http://psychtoolbox.org/
LIBSVM	Chang and Lin[2]	https://www.csie.ntu.edu.tw/~cjlin/libsvm/
Other		
Behnke-Fried depth electrodes	AD-TECH Medical Instrument Corp.	https://adtechmedical.com/depth-electrodes
ATLAS neurophysiology system	Neuralynx Inc.	https://neuralynx.com/news/techtips/atlas-neurophysiology-system-for-cogneuro-applications

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Florian Mormann (florian.mormann@ukbonn.de).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data and custom-built MATLAB code can be found in a GitHub repository (https://github.com/EstherKutter/Neuronal-Codes-For-Arithmetic-Rule-Processing-In-The-Human-Brain).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Nine neurosurgical patients (4 male, all right-handed, mean age 43.3 years) undergoing treatment for pharmacologically intractable epilepsy participated in the study. Informed written consent was obtained from each patient. All studies conformed to the guidelines of the Medical Institutional Review Board at the University of Bonn, Germany. Other parts of the current data set were published in a previous publication.[22]

METHOD DETAILS

Experimental Task and Stimuli

Subjects performed a calculation task that required them to calculate the result of a simple arithmetic problem (Figure 1A). During experimental sessions, subjects sat in bed, facing a touch-screen laptop (display diagonal 11.7 in, resolution 1366x768 px) on which stimuli were presented at a distance of approximately 50 cm. They were not informed about hypotheses or purposes of the experiment, in order to avoid any bias.

Each trial began after a 500 ms fixation phase. Stimuli were presented successively in the order operand 1 – operator – operand 2, for 500 ms each, followed each by a 800 ms delay phase. Afterwards, subjects responded in a self-paced manner by touching the number matching the result of the calculation on a number pad showing the arabic numerals 0 to 9 that was presented on the screen. After a 500 ms feedback display ('richtig' [correct] or 'falsch' [false]) the next trial was started automatically.

All stimuli were presented within a filled gray circle (diameter approx. 6° of visual angle) on a black background. During fixation and delay phases, we presented a white fixation spot in the center of the gray area. During stimulus presentation, the fixation spot disappeared to avoid confusion with nonsymbolic stimuli and to distinguish it clearly from the nonsymbolic 'zero'-stimulus that was included as a potential operand 2-stimulus for control purposes.

Numerical values of operand 1 ranged from 1 to 5, and were in two visual 'formats', either 'nonsymbolic' arrays of randomly placed black dots of varying sizes with the number of dots corresponding to the respective numerical value ('numerosities'), or 'symbolic' black Arabic digits at randomized locations ('numerals'). Number stimuli of operand 2 ranged from 0 to 5, and were the same as for operand 1. The nonsymbolic 'zero'-stimulus was presented as the empty gray circle without fixation spot.

We used two 'protocols', standard and control displays, for both nonsymbolic and symbolic number formats (Figure 1B) in order to control for low-level visual features. The standard nonsymbolic numerosity displays consisted of dots at randomized locations and of pseudo-randomly varied sizes (diameter 0.3° to 0.8° of visual angle); in the control displays, we equated the overall screen area and density of the dots across numerosities. For the Arabic numerals, different fonts were used as standard (Helvetica, 34 pt) and control (DS-Digital, 34 pt) displays. A session comprised 50% nonsymbolic and 50% symbolic stimuli. Within each format, standard and control protocols were shown with equal probability of 50%.

We applied two different mathematical rules, i.e., addition and subtraction (Figure 1C). Two distinct cues, i.e., the mathematical sign ('+' or '−') or a verbal analogue ('und' [add] and 'weniger' [subtract]), were used for each rule (all Helvetica, 34 pt, presented in the center), in order to dissociate neuronal activity related purely to visual properties of the operator from the rule that it represented.

Overall, the task comprised seven factors. Five of these factors were varied systematically: Format (symbolic vs. nonsymbolic), protocol (standard vs. control) and numerical value of operand 1 (1–5), as well as mathematical rule (addition vs. subtraction) and rule cue (sign vs. word). Operand 2 was always of the same format and protocol as operand 1, but with random numerical values 0–5, albeit guaranteeing calculation results between 0 and 9. Due to this constraint, it was impossible to balance the other two factors 'numerical value of operand 2' (e.g., '5' is less likely to appear than '1', given that 'X–5' is only valid for X = {5}, but 'X-1' is valid for X = {1,2,3,4,5}), and 'numerical value of calculation result' (e.g., '4' is more likely to appear than '9', given the possible combinations of operands to obtain this result).

Each session consisted of a total of 320 trials and was divided into four blocks of 80 trials each, comprising the different conditions in pseudo-random order. To familiarize subjects with the task, sessions started with 10 rehearsal trials that were excluded from further analysis.

Neurophysiological Recording

To localize the epileptic focus for possible clinical resection, each subject was implanted bilaterally with chronic intracerebral depth electrodes in the medial temporal lobe (MTL). The exact electrode locations and numbers were defined exclusively by clinical criteria and varied across subjects. We used 9–10 clinical Behnke-Fried depth electrodes (AD-Tech Medical Instrument Corp., Racine, WI) to record neuronal signals. Each depth electrode contained a bundle of nine platinum-iridium micro-electrodes protruding from its tip by approximately 4 mm. Each bundle consisted of eight high-impedance active recording channels and one low-impedance reference electrode. A 256-channel ATLAS neurophysiology system (Neuralynx Inc., Bozeman, MT) was used to filter (bandwidth 0.1–9,000 Hz), amplify and digitize (sampling rate 32768 Hz) the differential neuronal signals (recording range ±3200 μV). The Cheetah software (Neuralynx Inc., Bozeman, MT) was used to synchronize the behavioral data with the recorded spikes via 8-bit timestamps.

Neuronal signals were band-pass filtered (bandwidth 300–3,000 Hz), then spikes were detected and pre-sorted automatically using the Combinato software.[54] Manual verification and classification as artifact, multi- or single unit was based on spike shape and its variance, inter-spike interval distribution per cluster and the presence of a plausible refractory period. Only units that responded with an average discharge rate of >1 Hz during stimulus presentation (fixation onset to delay 2 offset) were included in the analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Neuronal Analysis of Variance (ANOVA)

Only single units (n = 585) were included in the following analyses. All analyses were performed separately for each MTL area to identify regional differences (PHC: 126 units; EC: 107 units; HIPP: 199 units; AMY: 153 units). As all participants performed the task with high proficiency (98.5% ± 0.6%, range 90.3%–99.8%), we did not exclude the negligible number of error trials from the analyses.

For each unit, activity was analyzed separately for the different task phases involving rule processing. For each stimulus phase (calculation rule and operand 2 phase), discharge rates were measured in a 400 ms window starting 200 ms after stimulus onset. For each delay phase (rule delay and delay 2 phase), activity was assessed in a 700 ms window starting 200 ms after delay onset (latency chosen based on Mormann et al.[55] In total, six factors were analyzed: 'mathematical rule' (addition/ subtraction) and 'rule cue' (word/ sign), as well as 'format' (nonsymbolic/ symbolic), 'numerical value of operand 1' (numbers 1–5), 'numerical value of operand 2' (numbers 0–5), and 'numerical value of calculation result' (numbers 0–9). We pooled over the factor 'protocol' given its incomparability for the different formats[4] and its irrelevance for the processing of the rule cues. For each task phase, we performed an ANOVA considering only those factors relevant for that phase. That is, for the calculation rule and rule delay phase, a 4-way ANOVA with the factors 'mathematical rule', 'rule cue', 'format' and 'numerical value of operand 1' was performed. For the operand 2 and delay 2 phase, we calculated a 6-way ANOVA with the factors 'mathematical rule', 'rule cue', 'format', 'numerical value of operand 1', 'numerical value of operand 2' and 'numerical value of calculation result'. All ANOVAs were evaluated at α = 0.01. A unit was counted as exclusively rule-selective ("Only Rule") if a significant main effect was observed for the factor 'mathematical rule', and there was no significant main effect for any other factor. Exclusive cue-selectivity ("Only Cue") was defined analogously.

To evaluate the significance of unit proportions, we subjected the number of significant neurons to a binomial test with an *a priori* probability of $p = 0.01$ corresponding to the alpha level for neurons to be regarded as significant, Bonferroni-corrected for multiple comparisons across different areas ($n = 4$).

Support Vector Machine (SVM) Classification

All single units were included in the following population analyses, irrespective of any selectivities found in the ANOVA. For each unit, data were divided into two classes, assigning the label 'addition' to trials with the cues 'und' [add] and '+', and the label 'subtraction' to the trials with the cues 'weniger' [subtract] and '–'. With 80 trials per cue, each class comprised 160 trials. For temporal cross-training classification, spike trains of each unit were smoothed (Gaussian kernel, $\sigma = 150$ ms, window size 300 ms) trial-wise within the trial window of 0–4500 ms (i.e., from fixation onset to 100 ms after delay 2 offset). An SVM classifier with a default linear SVM kernel[56] was then trained on the instantaneous firing rates at a certain time point, and tested on firing rates at different time points (sampling interval 50 ms).

We applied 10-fold cross-validation, i.e., we created 10 equal-size complementary splits of our dataset, balancing conditions within each split. Then, 9 splits were used as training set (comprising 288 trials), the remaining split was used as test set (comprising 32 trials). All firing rates were normalized by z-scoring (mean and standard deviation obtained from training data only), then we fitted the classifier to the training data and assessed the predictive accuracy by counting the instances that a certain activity pattern of the test data was labeled correctly. This process was repeated 10 times, using each of the 10 splits exactly once as the validation set. The results were then averaged across all splits.

To identify temporal clusters during which accuracy differed significantly from chance level (50% for two classes), the analysis was repeated with randomly shuffled trial labels ($n_{perm} = 1000$), and a cluster permutation test[57] was performed. In short, we identified temporal clusters of statistical significance by comparing the true accuracy values against the distribution of random ones ($\alpha_{clus} = 0.05$). The significance of these 'candidate clusters' was then evaluated by comparison with the clusters of the random data ($p_{rank} < 5\%$), using cluster size as a test statistic (i.e., number of connected significant 'pixels' in the cross-temporal accuracy matrix, or cluster length for the 'diagonal curve' when training and test time points were identical).

High accuracy values do not imply *per se* that the classifier has learned to encode abstract rule information; comparable accuracies might also be achieved if the SVM had learned to encode only one specific cue perfectly, but remained at chance level for the other three cues. To account for this, an SVM (with the same settings as above) was trained and tested on the firing rates obtained by averaging across the significant interval when training and testing at the same time point. In this fixed-window analysis we used the following time windows: PHC: 1950–3000 ms and 3150–4000 ms; EC: 2200–3200 ms; HIPP: 2250–4400 ms; AMY: 2700–3600 ms. We generated a confusion matrix which counted the frequency at which a trial of a certain rule cue was assigned different labels by the classifier, and calculated the accuracy per mathematical rule by averaging classification probabilities across the corresponding cues. In PHC, we trained a classifier and assembled confusion matrix and classification probabilities separately for each of the two significant time windows. Then, we averaged across both models to obtain one overall confusion matrix and overall average accuracies per rule. To evaluate significance, we repeated the analysis with shuffled labels ($n_{perm} = 1000$) and applied a permutation test ($\alpha = 0.05$).

As control, we equalized population sizes by drawing a random subset of units per area ($n = 107$) and re-calculated all analyses. This process was repeated 10 times, and the overall statistic was taken to be the mean of the stratified populations.

Finally, we assessed the units' ability to distinguish the two cue types (as opposed to the arithmetic rule information) by assigning the label 'word' to the 'und' [add] and 'weniger' [subtract] trials, and the label 'sign' to the '+' and '–' trials. We then repeated the temporal cross-training classification analysis and trained an SVM classifier on the window significant in the permutation test to generate the confusion matrix and average accuracy per cue type. The same procedures and settings as above (except for the labeling) were used for this control analysis.

Generalization of SVM Classification across Rule Cue Notations

To assess how well the results of the SVM classification might generalize to a different cue type, spike trains of all units were again trial-wise smoothed within the trial window (parameters as above), and labeled as before. Data were then divided into a training and a test set according to the rule cue.

First, all word trials (i.e., 'und' [add] trials labelled 'addition' and 'weniger' [subtract] trials labelled 'subtraction') served as training dataset. We applied 10-fold cross-validation, i.e., we split the training data into 10 balanced subsamples and used 9 splits as training dataset (comprising 144 trials). All sign trials (i.e., '+' trials labelled 'addition' and '–' trials labelled 'subtraction') served as test dataset (comprising 160 trials). Temporal cross-training classification was then performed using the same parameters and procedures as before. This process was repeated 10 times, leaving out each of the 10 subsamples exactly once. The results were then averaged across all splits. Again, significant temporal intervals were identified using a cluster permutation test ($n_{perm} = 1000$; $\alpha_{clus} = 0.05$; $p_{rank} < 0.05$). Generalization was analyzed also in the opposite direction, i.e., using sign trials as training dataset and word trials as test dataset, following the same procedures and settings as above.

Next, we identified synchronous intervals, i.e., time windows for which significant classification was observed for both directions of generalization. Intervals in which significance breaks of at most 150 ms for either one of the test directions occurred were

considered synchronous. Based on this criterion, we identified the following time windows: PHC: 2450–2650 ms and 3250–3950 ms; HIPP: 2950–4250 ms. For each direction, an SVM classifier (with the same settings as above) was trained on the firing rates obtained by averaging the training data across these synchronous intervals. Then, we tested the models on the firing rates obtained by averaging the test data across the same time window, and generated the confusion matrix. As before, in PHC, a classifier was trained for each of the two time windows. We then averaged the confusion matrices obtained for each interval to get one overall confusion matrix.

Generalization was then judged successful if (a) we found synchronous intervals of significant classification in the temporal cross-training analysis, and (b) performance in the fixed-window analysis was significant in a permutation test (n_{perm} = 1000; α = 0.05) for each arithmetic rule after averaging classification accuracies across both directions of generalization.

Current Biology, Volume *32*

Supplemental Information

Neuronal codes for arithmetic rule processing in the human brain

Esther F. Kutter, Jan Boström, Christian E. Elger, Andreas Nieder, and Florian Mormann

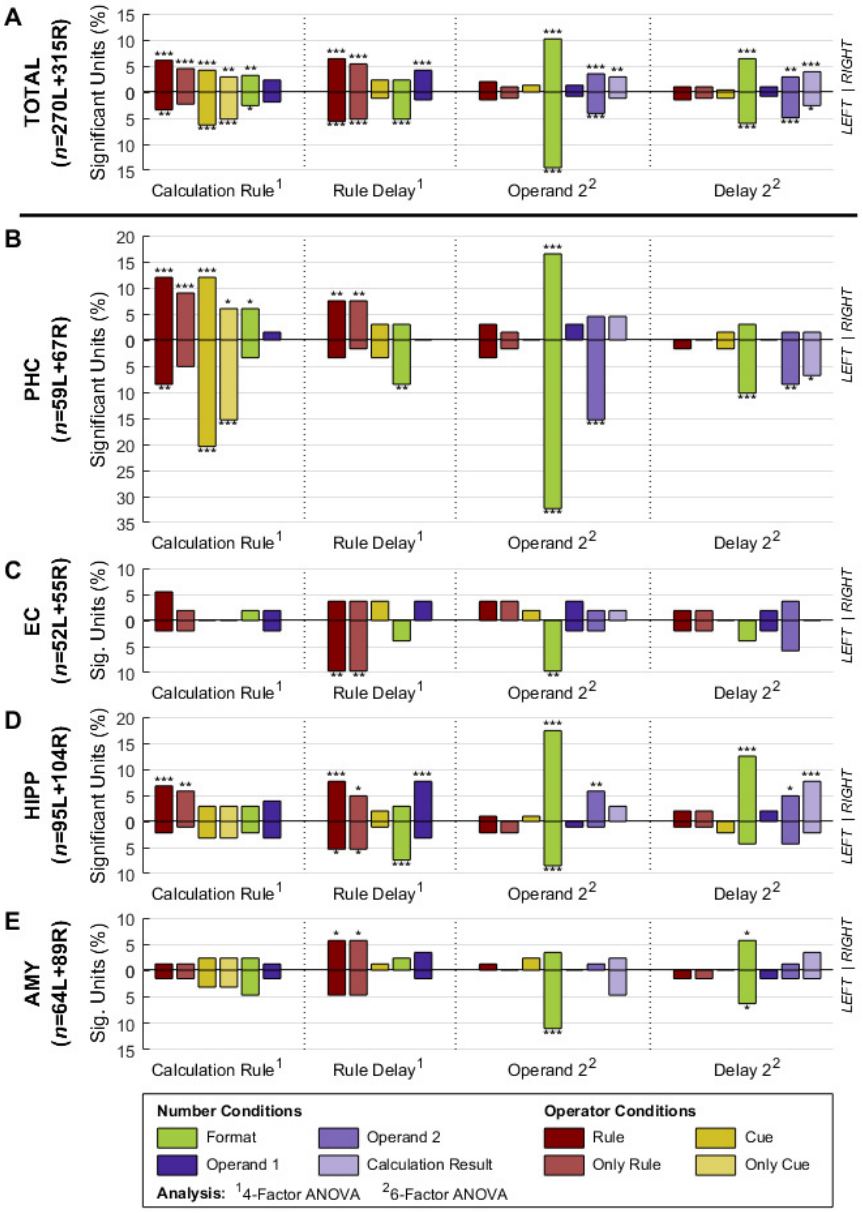


Figure S1: Neuronal Selectivity of MTL Single Units per Hemisphere. Related to Figure 3. Proportions of single units significant to different task factors for different MTL regions and hemispheres: **(A)** total population, **(B)** parahippocampal cortex, **(C)** entorhinal cortex, **(D)** hippocampus, and **(E)** amygdala. Proportions of units from the right and left hemisphere are depicted in the upper and lower rows, respectively. ANOVAs for the different task phases were evaluated at $\alpha = 0.01$. Neurons with an effect for 'arithmetic rule', but no concurrent other main effects are termed 'exclusively rule-selective' ("Only Rule"); same for factor 'rule cue'. Numbers of significant neurons were subjected to a binomial test, Bonferroni-corrected for multiple comparisons across areas ($n_1 = 4$) and hemispheres ($n_2 = 2$); asterisks indicate significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

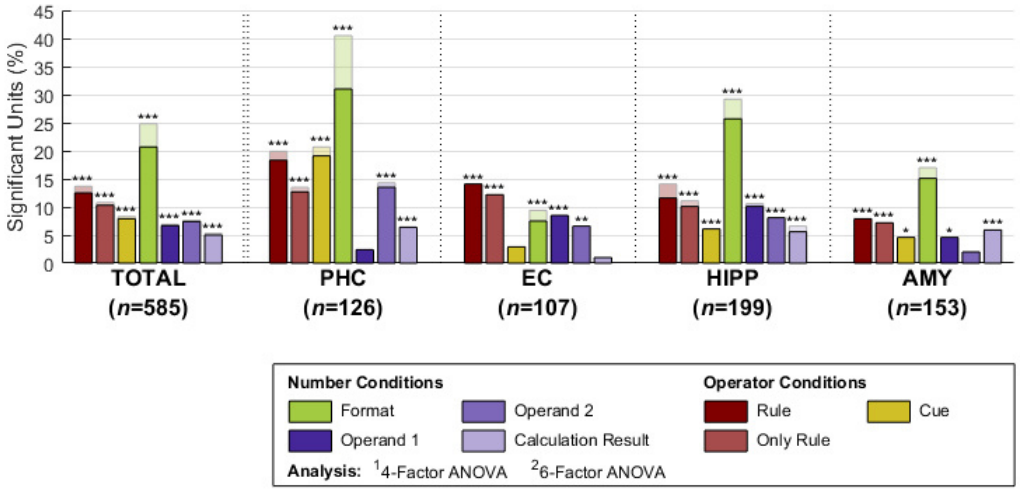


Figure S2: Neuronal Selectivity of MTL Single Units *a*cross Periods. Related to Figure 3. Proportions of single units significant in any (i.e., one or more) of the task periods (solid bars) along with percentages added up across *the four* trial periods (light bars). ANOVAs for the different task phases were evaluated at $\alpha = 0.01$. Neurons with an effect for 'arithmetic rule', but no concurrent other main effects are termed 'exclusively rule-selective' ("Only Rule"). Numbers of significant neurons were subjected to a binomial test, Bonferroni-corrected for multiple comparisons across areas ($n_1 = 4$) and task phases ($n_2 = 4$); asterisks indicate significance (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

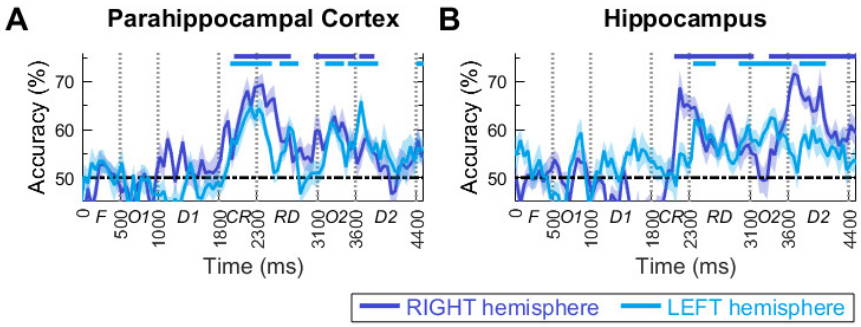


Figure S3: Rule Decoding in Different Hemispheres using an SVM Classifier. Related to Figure 4. Classification accuracy for decoding arithmetic rule information in **(A)** parahippocampal cortex: 67 units recorded from the right hemisphere, 59 units from the left hemisphere, and **(B)** hippocampus: 104 units recorded from the right hemisphere, 95 units from the left hemisphere. SVM classifiers were trained on the instantaneous firing rates across the trial period. The dashed line represents chance level (50 % for two classes). Light and dark blue bars above the data indicate significance ($p < 0.05$) in a permutation test for each hemisphere. Abbreviations at the axes indicate task phases: *F*, fixation; *O1*, operand 1; *D1*, delay 1; *CR*, calculation rule; *RD*, rule delay; *O2*, operand 2; *D2*, delay 2.

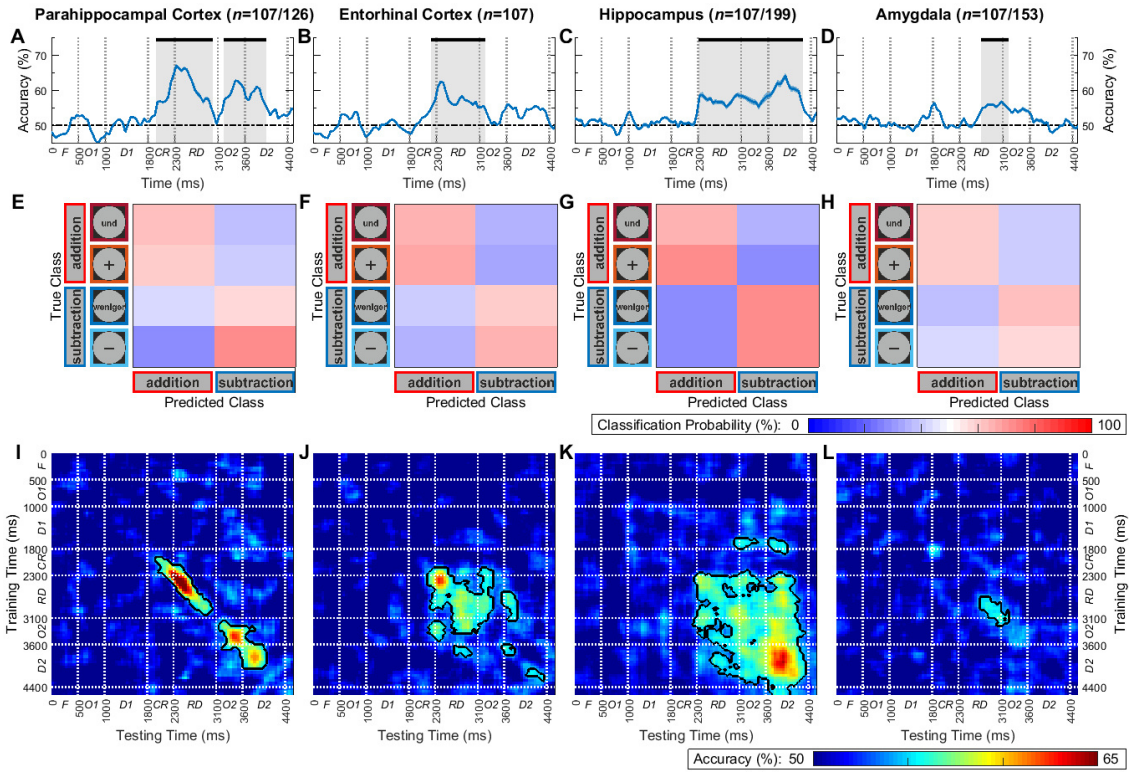


Figure S4: Rule Decoding in Sample-Equalized MTL Populations using an SVM Classifier. Related to Figure 4. Decoding performance when using random subsets of neurons per area, equalizing population size across all MTL regions (columns). **(A–D)** Average classification accuracy for decoding arithmetic rule information when training an SVM on the instantaneous firing rates across the trial period. The dashed line represents chance level (50 % for two classes). Black bars above the data and gray shaded areas indicate significance ($p < 0.05$) when testing against performance for SVMs trained on shuffled data in a permutation test. Abbreviations at the axes indicate task phases: *F,* fixation; *O1,* operand 1; *D1,* delay 1; *CR,* calculation rule; *RD,* rule delay; *O2,* operand 2; *D2,* delay 2. **(E–H)** Confusion matrix derived when training an SVM on firing rates averaged across the significant time window in (A–D), respectively. **E** shows the average of the confusion matrices obtained for each significant window (depicted in A). **(I–L)** Accuracy when training an SVM at a given time point of the trial and testing on another one (the main diagonals of the matrices correspond to the curves in A–D). Black contours indicate significance ($p < 0.05$) in a permutation test.

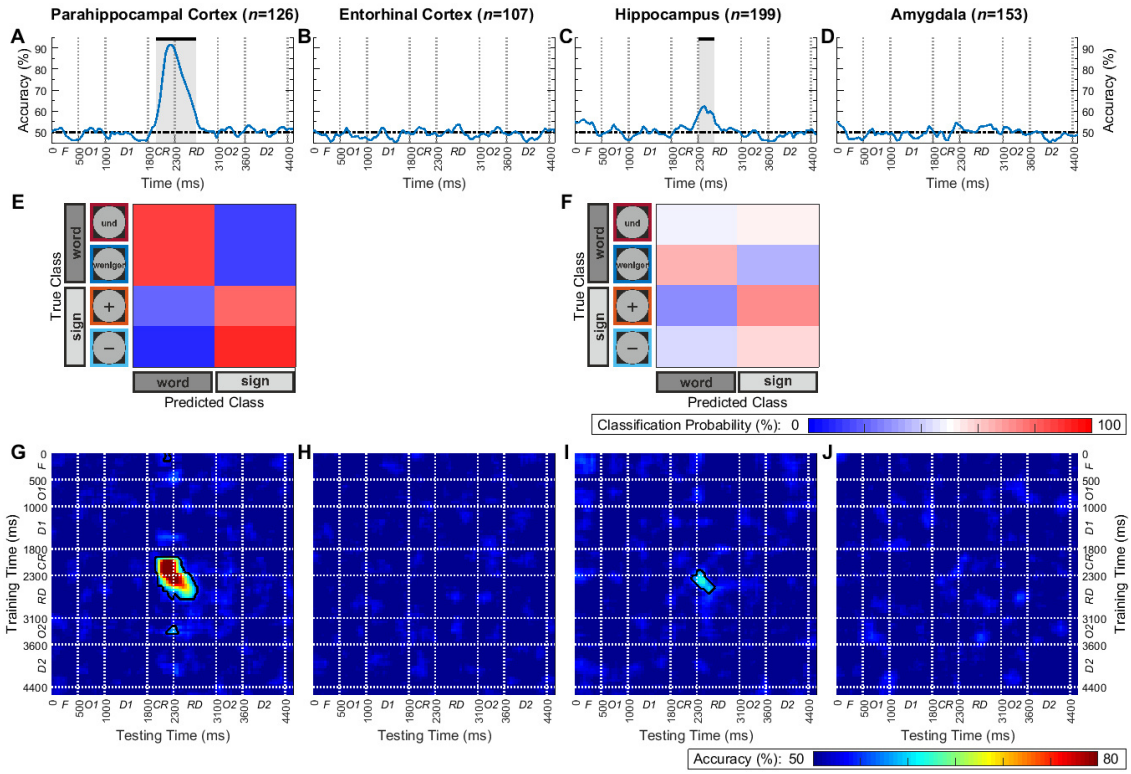


Figure S5: Rule Notation Decoding using an SVM Classifier. Related to Figure 4. Classifier performance for the four different MTL regions (columns). **(A–D)** Classification accuracy for decoding cue information when training an SVM classifier on the instantaneous firing rates across the trial period. The dashed line represents chance level (50 % for two classes). Black bars above the data indicate significance ($p < 0.05$) when testing against performance for SVM classifiers trained on shuffled data in a permutation test. Abbreviations at the axes indicate task phases: *F,* fixation; *O1,* operand 1; *D1,* delay 1; *CR,* calculation rule; *RD,* rule delay; *O2,* operand 2; *D2,* delay 2. **(E,F)** Confusion matrix derived when training an SVM on firing rates averaged across the significant time window in A and C, respectively. Significance was reached only in the PHC. **(G–J)** Accuracy when training an SVM classifier at a given time point of the trial and testing on another one (the main diagonals of the matrices correspond to the curves in A–D). Black contours indicate significance ($p < 0.05$) in a permutation test.

Publication 3: Distinct Neuronal Representation of Small and Large Numbers in the Human Medial Temporal Lobe

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**: 1998–2007. DOI: 10.1038/s41562-023-01709-3

nature human behaviour

Article

https://doi.org/10.1038/s41562-023-01709-3

Distinct neuronal representation of small and large numbers in the human medial temporal lobe

Received: 12 February 2023

Accepted: 31 August 2023

Published online: 02 October 2023

Check for updates

Esther F. Kutter[1,2], Gert Dehnen[1], Valeri Borger[3], Rainer Surges[1], Florian Mormann[1,4] ✉ & Andreas Nieder[2,4] ✉

Whether small numerical quantities are represented by a special subitizing system that is distinct from a large-number estimation system has been debated for over a century. Here we show that two separate neural mechanisms underlie the representation of small and large numbers. We performed single neuron recordings in the medial temporal lobe of neurosurgical patients judging numbers. We found a boundary in neuronal coding around number 4 that correlates with the behavioural transition from subitizing to estimation. In the subitizing range, neurons showed superior tuning selectivity accompanied by suppression effects suggestive of surround inhibition as a selectivity-increasing mechanism. In contrast, tuning selectivity decreased with increasing numbers beyond 4, characterizing a ratio-dependent number estimation system. The two systems with the coding boundary separating them were also indicated using decoding and clustering analyses. The identified small-number subitizing system could be linked to attention and working memory that show comparable capacity limitations.

When asked to judge the number of briefly presented items in a set, humans show a behavioural dichotomy[1]. Participants respond fast and accurately for small numbers up to about 4 in a process termed 'subitizing'[2]. However, for larger numbers beyond 4, participants show increasingly slower and more imprecise number 'estimation' that is dependent on the ratio between the numbers to be compared[2–5].

On the basis of behavioural measures, it has been argued that the observed judgement differences arise from one and the same estimation system whose negligible ratio-dependent imprecision for small numbers gives rise to a seeming dichotomy in underlying mechanisms[6,7]. Others, in contrast, maintain that subitizing and estimation reflect two distinct mechanisms for assessing small versus large numbers[2–5]. Explorations into underlying brain mechanisms using blood flow imaging or electroencephalography remained similarly inconclusive; while some studies argue for a single underlying mechanism[8–12], others propose two separable number systems[13–15].

In this Article, to address this century-old debate about a single or two distinct mechanisms for number representations, we recorded single-neuron activity in the medial temporal lobe (MTL) of neurosurgical patients who judged numerical quantity[16,17]. If small and large numbers are represented by the same neuronal mechanism, a continuous code across small and large numbers is anticipated. However, if small and large numerosities engage distinct mechanisms, two different coding schemes with a discontinuity reflecting the change from one mechanism to the other is expected.

Results

We asked 17 human participants to quickly judge the parity (even versus odd) of numbers ranging from 0 to 9 shown as dot arrays on a computer screen. The simple parity task is suited to test a broad range of explicit number representations devoid of other cognitive factors (such as working memory), and in short time for the participants. In each trial, a numerosity

[1]Department of Epileptology, University of Bonn Medical Center, Bonn, Germany. [2]Animal Physiology, Institute of Neurobiology, University of Tübingen, Tübingen, Germany. [3]Department of Neurosurgery, University of Bonn Medical Center, Bonn, Germany. [4]These authors contributed equally: Florian Mormann, Andreas Nieder. ✉e-mail: florian.mormann@ukbonn.de; andreas.nieder@uni-tuebingen.de

https://doi.org/10.1038/s41562-023-01709-3

was flashed for 500 ms after a short fixation period, followed by a brief delay during which the number stimulus was removed (Fig. 1a). Afterwards, participants decided whether the number of dots had been even or odd by pressing the left or right arrow key, respectively, on the keyboard as indicated on the response screen. The keys associated with the respective response were switched between blocks to control for potential motor bias. Different stimulus protocols were used to control for non-numerical visual parameters: dots were shown in a standard (variable dot size and arrangement) and two control displays (constant total dot area and dot density across numerosities, and linear arrangement) (Fig. 1b). The numerosity and protocol of the stimuli varied randomly from trial to trial.

Behaviour

The participants' performance showed well-known behavioural effects indicative of two different representational systems. Small countable numerosities from 1 to 4 were equally effortlessly judged with only few errors (Fig. 1c) and short reaction times (RTs) (Fig. 1d), as expected for subitizing[2–4]. In contrast, numbers 5 and higher were judged with noticeably increasing error rates and RTs indicative of number estimation. This observation was bolstered by calculating the discontinuity point that signals a change in the slope[18], which could be determined for 14 of the 17 participants. We found average discontinuity points of 3.7 and 3.6 for error rates and RTs, respectively, as the upper boundary of the subitizing range (Fig. 1c,d). The errors seen for numbers larger than 5 argue that participants were not symbolically counting items as serial counting would be error-free and moreover has been shown to be impossible in afterimages[19,20]. Asymmetric switch cost effects for the transition from subitizing to estimation versus the transition from estimation to subitizing were not observed[21,22] ($P_{\text{switch condition}} = 0.88$; two-factor analysis of variance (ANOVA) with factors 'numerical value' (0–9) × 'switch condition' (switch versus non-switch)). Consistent with previous reports[23,24], the empty set (number zero) elicited distinct behavioural effects due to its special status as a latecomer in number concepts[25].

Neuronal responses

To test the long-standing hypothesis of different enumeration systems for small versus large numbers, we recorded action potentials of 801 single neurons in the MTL of the 17 participants while they performed the number task. Many neurons were activated in a tuned fashion to the numerical value of the sample stimulus. They responded strongest to their respective preferred numerosities and decreased their activity progressively with increasing numerical distance (Fig. 2a–d). We statistically identified number-selective neurons by applying a sliding-window analysis to all cells[16]. We combined a two-factor ANOVA with factors 'numerical value' (0–9) × 'protocol' (standard versus control) to detect tuning to numerical values, and a separate Mann–Whitney *U* test with factor 'parity' (even versus odd) to exclude neurons responsive to parity judgements (both evaluated at $\alpha = 0.01$). Across all four areas individually, a substantial proportion of neurons showed a significant main effect for the factor 'number' ($P < 0.001$; binomial test with $P_{\text{chance}} = 0.01$), but no effect for the factors 'protocol' or 'parity' (Fig. 3a). Across all four areas combined, 15.1% of MTL neurons (121/801) showed an exclusive significant main effect for the factor 'number' (Fig. 3a). Each of the tested numerosities (0–9) constituted the preferred numerosity of individual selective neurons (Fig. 3b); differences in these proportions were not due to response preferences for specific numerosities, but consistent with random variation ($P = 0.15$; Mantel–Haenszel test). Similarly, response latencies across the four MTL regions (parahippocampal cortex (PHC), entorhinal cortex (EC), hippocampus (HIPP) and amygdala (AMY)) did not reveal significant differences ($P = 0.87$; Kruskal–Wallis test).

Neuronal tuning characteristics

To explore hypothesized different physiological mechanisms for the representation of small and large numerosities, we first analysed the tuning curves of number-selective neurons. We calculated the numerosity tuning functions of all numerosity-selective neurons using standardized activity (*z*-score relative to baseline activity) (Fig. 3c). Apart from the well-known number-tuning characteristics, that is, maximum activity to the preferred number and progressively reduced firing rates as distance from the preferred number increased, several distinguishing features in the tuning to small versus large numbers emerged.

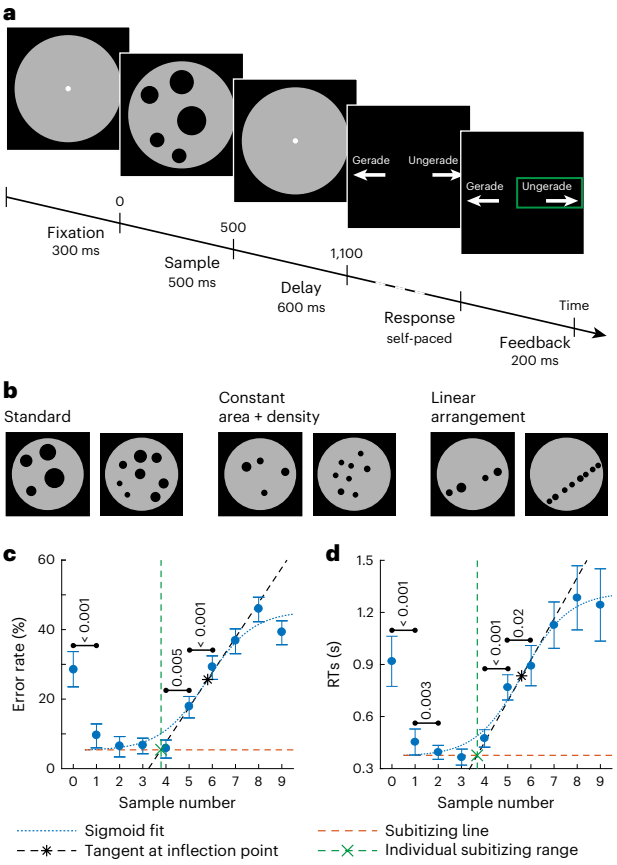


Fig. 1 | Behavioural task, stimuli and behavioural performance. a, Parity judgement task. Participants were required to indicate whether the number of dots was even ('gerade') or odd ('ungerade') by pressing the left or right arrow key, respectively (or vice versa). **b**, Sample number protocols. Dot arrays represented numerosity. They were shown in a standard layout with variable dot size and position (left), in a control layout with equalized total area and density of the dots (middle), and additionally as linearly arranged dots (right). Numerical values covered the range 0–9; exemplary dot displays for numbers 4 and 8 for each protocol are shown. **c**, Behavioural performance. Mean error rates and error bars denoting standard error of the mean (s.e.m.) are shown ($n = 17$). Values above small horizontal bars indicate *P* values for pair-wise comparisons (two-sided, paired *t*-test, Bonferroni-corrected for multiple comparisons of numbers ($n = 9$)); all other pair-wise comparisons were not different ($P > 0.05$). The subitizing boundary (green dashed line) is defined as the intersection point of the tangent (black dashed line) at the inflection point (black star) of a sigmoid fit (blue dotted line) to the error rates (excluding zero), and the subitizing line (red dashed line) at which the sigmoid curve intersects the *y* axis. **d**, RTs. Median and error bars denoting s.e.m. are shown ($n = 17$). Values above small horizontal bars indicate *P* values for pair-wise comparisons (two-sided, Wilcoxon signed rank test, Bonferroni-corrected for multiple comparisons of numbers ($n = 9$)); all other pair-wise comparisons were not different ($P > 0.05$). Conventions for the subitizing boundary as in **c**.

https://doi.org/10.1038/s41562-023-01709-3

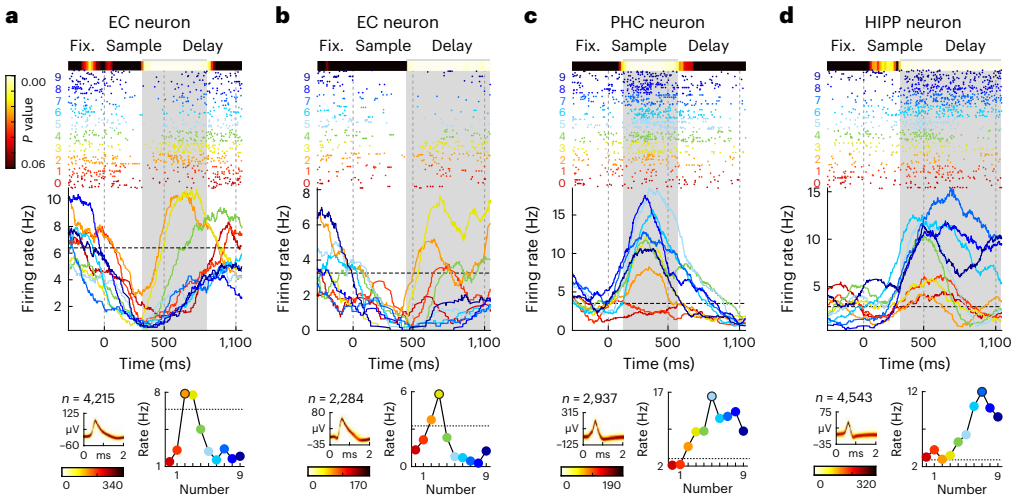


Fig. 2 | Responses of number-selective neurons. a, Example neuron from EC tuned to small number 2. Top: dot-raster histogram. Each row indicates one trial (colours correspond to presented numbers); each dot represents one action potential. Middle: corresponding mean instantaneous firing rates across trial time obtained by averaging responses to each number (smoothed using a 150 ms Gaussian kernel). Colours correspond to sample number. The horizontal dotted line depicts spontaneous activity (average across fixation periods). The grey shaded area represents the significant number-selective interval according to the sliding-window ANOVA (colour-coded *P* values above each panel). Bottom (left): density plot of the recorded action potentials, colour darkness indicating the number of overlapping wave forms according to the colour scale at the bottom. Bottom (right): number tuning function (average firing rate in the selective trial interval plotted against sample number). The horizontal dotted line indicates spontaneous firing rate. **b**, Example neuron from EC tuned to small number 3. Same layout as in **b**. **c**, Example neuron from PHC tuned to large number 5. Same layout as in **b**. **d**, Example neuron from HIPP tuned to large number 7. Same layout as in **b**.

First, tuning functions to small preferred numerosities 0–3 showed systematic surround suppression below spontaneous activity to non-preferred numbers, whereas tuning functions to large numerosities 4–9 returned to spontaneous activity for non-preferred numbers (Fig. 3c; see also example neuron tuning functions in Fig. 2a–d). We compared the firing rates elicited by non-preferred numbers (that is, at the flanks of the tuning curves) to the neurons' baseline activity. We found that firing rates at the flanks were significantly smaller than baseline activity in neurons tuned to each of the preferred numbers 0 to 3 (*P* values numerosity 0: 1.4×10^{-15}; numerosity 1: 0.0024; numerosity 2: 8.7×10^{-14}; numerosity 3: 0.0086; one-sided Wilcoxon signed rank tests), but not different from baseline in neurons tuned to each of the preferred numbers 4 to 9 (all *P* values >0.98; one-sided Wilcoxon signed rank tests) (Fig. 3d). The sharp cut in surround suppression between 3 and 4 was not due to tuning functions for preferred numbers larger than 3 becoming too wide to detect suppression. This argues for a physiological effect rather than a tuning-function resolution issue.

Second, and correlating with this tuning-flank suppression, we observed systematic differences in the amplitudes of the tuning curves. We fitted Gauss functions to the tuning curves and derived the amplitude value as a quantitative measure for the amplitude of the tuning functions[25]. Tuning curve amplitudes of neurons tuned to small numbers (0–3) were significantly smaller compared to large number (4–9) tuning curves ($P < 0.001$; one-sided Mann–Whitney *U* test), whereas tuning amplitudes were indifferent within the groups of neurons tuned to small ($P = 0.38$; Kruskal–Wallis test) and large numbers ($P = 0.69$; Kruskal–Wallis test) (Fig. 3e).

Third, tuning selectivity showed a dichotomy between small and large numbers. Since small numbers in the subitizing range can be discriminated more accurately (Fig. 1c), and more accurate discrimination is linked to more selective (that is, narrower) tuning functions[26–28], systematic differences in number tuning selectivity between the subitizing versus estimation range are expected. Thus, we derived the sigma value from the Gauss fits to quantify tuning width[29]. The tuning widths for neurons tuned to numbers 0–3 were small and not different in value ($P = 0.9$; Kruskal–Wallis test). Note that sigma as a measure of tuning width can be much smaller than 1, which is why the stable tuning widths in the subitizing range are not due to a floor effect. Around preferred number 4 or 5, a turning point emerged with tuning widths systematically increasing in a linear fashion towards larger numbers, as expected for ratio-dependent estimation (Fig. 3f). The selectivity dichotomy of neurons across the range of numbers is in agreement with behavioural predictions and suggests separate mechanisms for the coding of small versus large numerosities.

To explore the categorically distinct representation of small versus large numerosities further, we performed a representational similarity analysis (RSA) by calculating the correlation coefficients of the *z*-scored firing rates between all pairs of numbers for number-selective neurons ($n = 121$). We hypothesized that neurons tuned to small numbers would show more similar firing rates to other small numbers and thus higher correlation coefficients within pairs of small numbers, whereas neurons tuned to large numbers would show higher correlation coefficients within pairs of other large numbers. The resulting matrix of correlation coefficient values suggests radically different coding for numerosity 0 (which was therefore excluded from this analysis), but also categorical differences between small and large countable numbers (Fig. 3g). We then quantified for which of the eight number boundaries (that is, 1|2, 2|3, …, 8|9) the difference between within- and across-category correlation values was most significant and thus best segregated these data into small versus large number representations. The highest and most significant correlation value difference between within- and across-categories ($r = 0.27$; $P = 1.12 \times 10^{-5}$) was found for the boundary 3 versus 4 (Mann–Whitney *U* test, Bonferroni-corrected for multiple comparisons of boundaries, $n = 8$) (Fig. 3h). This correlation analysis suggests categorically different encoding of small versus large numbers based on the selective neurons' firing rates. When applied to the entire set of single units regardless of numerosity selectivity (801 units), this analysis yielded qualitatively similar results (Supplementary Fig. 1).

https://doi.org/10.1038/s41562-023-01709-3

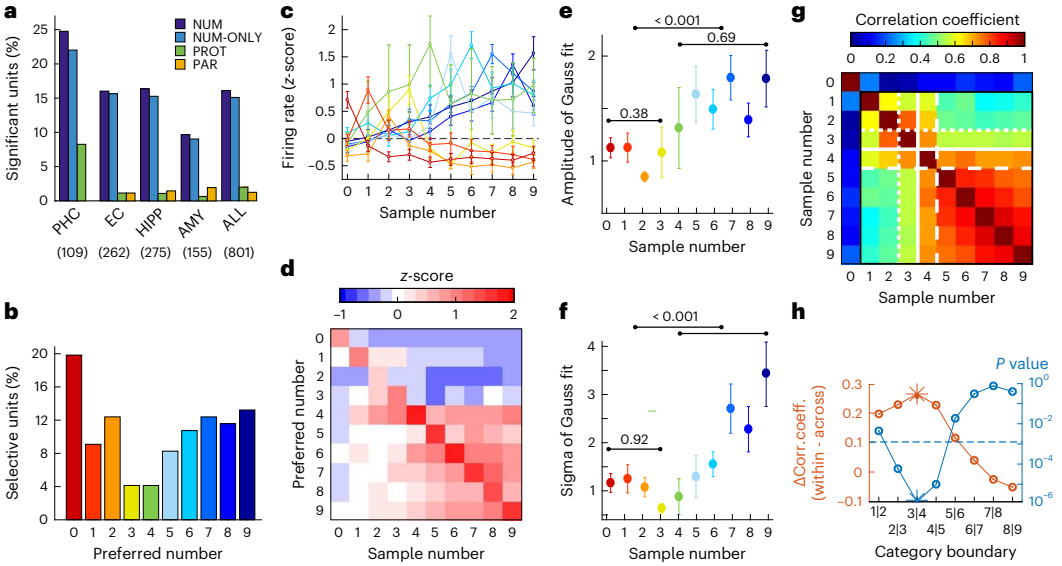


Fig. 3 | Tuning characteristics of number-selective neurons. a, Proportions of neurons with a significant main effect for 'number' (NUM), 'exclusively number' (NUM-ONLY), 'number protocol' (PROT) in a two-way ANOVA, or 'parity' (PAR) in a Mann–Whitney *U* test, evaluated at $\alpha = 0.01$, separately for each MTL region (number of recorded units in brackets). **b**, Proportion of neurons tuned to different preferred numbers. **c**, Average *z*-scored tuning curves of number-selective neurons tuned to the ten numbers (colour-coded as depicted in **b**). Error bars denote standard error of the mean (s.e.m.). **d**, Average (*z*-scored) firing rate per preferred number (rows) colour-coded relative to baseline activity. Blueish colours indicate suppression; reddish colours indicate enhancement of firing rates relative to baseline activity. **e**, Average tuning amplitude per preferred number derived from Gauss fits to tuning curves. Standard errors denote s.e.m. Amplitudes did not differ for units preferring small numerosities 0–3 ($n = 55$) and for units preferring large numerosities 4–9 ($n = 73$) (Kruskal–Wallis tests; $P > 0.05$, n.s.) but were significantly different between both groups, as indicated by the *P* value above the small horizontal bar (one-sided Mann–Whitney *U* test). **f**, Average tuning selectivity per preferred number as measured by sigma from Gauss fits to tuning curves. Error bars denote s.e.m. Sigma was small and constant for small numbers but increased in proportion with the value of large numbers. Sigmas did not differ for units preferring small numerosities 0–3 ($n = 55$), but were significantly different between both groups, as indicated by the *P* value above the small horizontal bar (one-sided Mann–Whitney *U* test). **g**, Correlation coefficients of the *z*-scored firing rates between pairs of numbers for all number-selective neurons ($n = 121$). Firing rates were more similar (higher correlations, corresponding to reddish colours) for numbers from the same number (small or large) category (upper-left and lower-right square), compared with responses for numbers from a different category (lower-left and upper-right square). White lines depict significant number category boundaries (solid line is most significant), dividing correlations into small versus large number categories. **h**, Evaluation of the goodness-of-fit of different number boundaries. Orange values depict the differences of correlation coefficients when segregating small versus large number categories (excluding zero) at different boundaries. The corresponding *P* values (two-sided Mann–Whitney *U* test) for these coefficient differences are shown in blue. Boundary 3 versus 4 (asterisks) divides the data most significantly into two number categories. The blue dotted line indicates $\alpha = 0.01$, Bonferroni-corrected for multiple comparisons ($n = 8$).

Population decoding using SVM classifiers

In addition to single neurons, neural populations carry information about neuronal computations[30]. Next, we therefore explored potential decoding discontinuities for numbers at the level of the population of selective neurons ($n = 121$) (Fig. 4a). Using a support vector machine (SVM) classifier, we first identified the time window of significant above-chance classification during a sliding-window classification analysis (60 ms to 1,200 ms after sample number onset; $\alpha = 0.01$) (Fig. 4b). Next, another classifier was trained with 50% of the data to discriminate the ten numbers, and then tested on the remaining 50% novel data from the same neuronal population and in the same time window to evaluate how well the model could decode each number based on information extracted from trials used during training. The classification probability of predicted numbers per truly presented number was then used to assemble a confusion matrix, with the main diagonal indicating correct labelling (Fig. 4c).

The classifier predominantly confused numbers from within the small-number category (upper-left square) and from within the large-number category (lower-right square), but not numbers across the small and large number category (in the lower-left and upper-right squares) (Fig. 4c). Zero was again excluded from the categorization analysis due to its distinctive difference relative to countable numbers. We quantified which of the eight number boundaries (that is, 1|2, 2|3, …, 8|9) resulted in the largest statistically significant differences in classification probabilities. A boundary between numbers 4 and 5 resulted in the largest and most significant difference in classification probability (difference 15%) between within- and across-categories ($P = 2.22 \times 10^{-6}$; Mann–Whitney *U* test, Bonferroni-corrected for multiple comparisons of boundaries, $n = 8$) (Fig. 4d). This population decoding analysis again indicates categorically different encoding of small versus large numbers, with a boundary between numbers 4 and 5. Again, this analysis yielded similar results when performed for the entire population of single units ($n = 801$; Supplementary Fig. 2).

Multi-dimensional state-space analysis and cluster analysis

Finally, to explore the dynamics of neuronal coding differences potentially pointing to two different number systems, we performed a multi-dimensional state-space analysis for the population of numerosity-selective neurons. At each point in trial time, the activity of *n* recorded neurons is defined by a point in *n*-dimensional space, with each dimension representing the activity of a single neuron ($n = 121$). The multi-dimensional space (used for quantitative analyses) is reduced to the three most informative dimensions for graphical depiction in three-dimensional (3D) state space. This results in 3D trajectories that are traversed for different neuronal states, that is, for the ten different numerical values (Fig. 5a). These trajectories reflect

https://doi.org/10.1038/s41562-023-01709-3

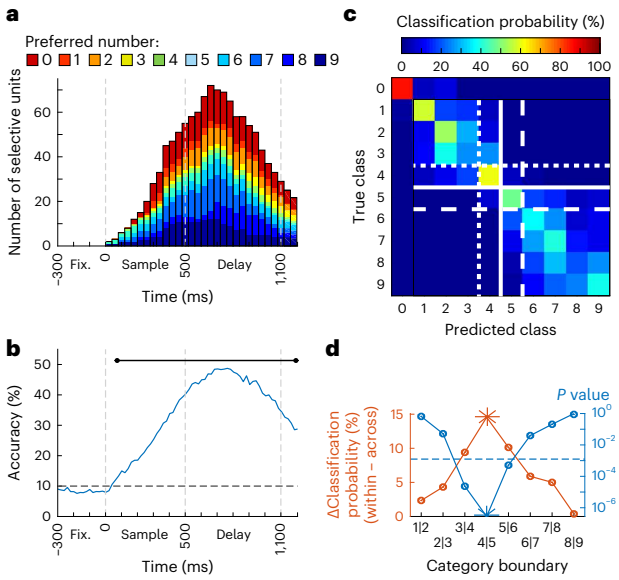


Fig. 4 | SVM classification analysis. a, Number of neurons that become selective to their respective preferred number during the course of the trial. **b**, Classification accuracy for decoding number information after training an SVM classifier on the instantaneous firing rates across the trial period for all number-selective neurons ($n = 121$). The dashed line represents chance level (10% for ten classes). The black bar above the data indicates significance ($P < 0.01$, one-sided permutation test compared with SVM trained on shuffled data). **c**, Confusion matrix derived from training an SVM classifier on firing rates averaged across the significant time window in the sliding-window analysis in **b** (60–1,200 ms). White lines depict the significant boundaries (highest significance for the solid, thick line) that divide the number range into small and large number categories. **d**, Evaluation of the goodness of different number boundaries. Orange values depict the difference in classification probabilities when segregating small versus large number categories (excluding zero) at different boundaries. The corresponding *P* values (two-sided Mann–Whitney *U* test) for these probability differences are shown in blue. Boundary 4 versus 5 (asterisks) divides the data most significantly into two number categories. The blue dotted line indicates $\alpha = 0.01$, Bonferroni-corrected for multiple comparisons ($n = 8$).

the instantaneous firing rates of the population of selective neurons as they evolve over time. Spatial closeness (that is, small distances) of the trajectories represents similarity in coding for the respective numerosities, whereas spatial disparity (that is, large distances) reflects coding dissimilarity.

The spatial layout of the trajectories evolving after sample onset until the end of the delay period again suggests two categorically different state spaces for numerosities 0–4 versus numerosities 5–9 (Fig. 5a). The trajectories representing numerosities 0–4 run in close vicinity to each other but (as expected) with increasing spatial gaps according to ordinal numerical distance. The same spatial pattern emerges within the group of trajectories representing numerosities 5–9. However, both trajectory categories are spatially segregated from each other by a large gap.

To statistically quantify this graphical grouping effect, we performed a cluster analysis on the neuronal population state space with averaged firing rates across the previously defined time window (60–1,200 ms after sample onset). The neural state space was then orthonormalized using principal component analysis. For visualization, only the first two dimensions (that is, PC1 and PC2) are shown (Fig. 5b). We first determined the optimal number of clusters for the data set by applying two measures: the Caliński–Harabasz index (also termed 'variance ratio criterion, VRC')[31], and the 'gap criterion' that determines the most dramatic decrease in error measurement (the 'elbow' or 'gap') of different cluster numbers[32]. Both measures indicated two clusters as the optimal cluster number for the dataset (Fig. 5c). We then applied unsupervised *k*-means clustering to partition all trials ($n = 160$) into two clusters[33]. The clustering algorithm detected one cluster that comprised the state spaces for numbers 0–4, and a second cluster consisting of state spaces for numbers 5–9 (Fig. 5d). Thus, the number state space is optimally described by two clusters that border between number representations 4 and 5. Again, performing this analysis for the entire population of single units ($n = 801$) yielded similar results (Supplementary Fig. 3).

Discussion

Our results provide evidence for two mechanisms encoding the continuous range of number. The number space from 0 to 9 was uninterruptedly covered by single neurons' overlapping tuning functions inherently ordered by number, and the activity of neuron populations was systematically arranged by numerical distances[16,34–36]. However, a coding dichotomy mirroring behavioural findings emerged within this representational continuum: neuronal tuning to small numbers in the subitizing range was more selective and ratio independent, whereas tuning widths increased in a ratio-dependent manner after a turning point around number 4. We also observed strong evidence for this coding dichotomy at the neuronal population level. This argues for a separate enumeration system for subitizing in addition to an estimation system[2–5]. Whether the current findings in the MTL transfer to other brain regions is currently not known and requires further exploration.

A defining feature of neuronal tuning in the subitizing range was surround suppression below baseline activity. Surround inhibition is a basic neuronal circuit operation[37,38] known to increase contrast sensitivity. Here, excitatory neurons firing in response to preferred stimuli recruit broadly tuned inhibitory interneurons that in turn suppress firing of neurons tuned to different preferred stimuli. Inhibition via interneurons is supposed to shape and sharpen the tuning to numerosities in the animal brain[39,40] and could mechanistically explain the more accurate number discrimination in the subitizing range. The time scale of surround inhibition to enable selective encoding in the subitizing range could be very fast. Moreover, the time delay of surround suppression with respect to classical receptive field excitation in the primate visual system has been reported to range from 15 to 60 ms (ref. 41), but (with a delay of 9 ms) can also act almost as suddenly as the direct-driving classical receptive field excitation signals[42]. Such short delays in surround inhibition are thought to emerge from a combination of feedforward, lateral and feedback connections to the target area[39,43]. While these mechanisms of surround suppression are a realistic assumption to explain the enhanced neuronal tuning in the subitizing range, they need direct testing in future experiments. With excitatory and inhibitory neurons identified in the human MTL[44–46], the necessary circuit components would readily be available to implement almost instantaneous surround inhibition for selective coding in the subitizing range.

Subitizing has been suggested to tap a different system in addition to that that for number estimation[5]. In contrast to number estimation, which is unaffected by attentional manipulations, subitizing is heavily dependent on attentive resources[47–49]. Attention-based processes that determine how many elements of information can be kept active in working memory have a very limited capacity of up to around four items[50,51], precisely the same set-size limit found for subitizing[52]. The mechanisms we discovered for subitizing may therefore well play a role for other capacity-limited processes, such as attention and working memory[49]. Similar to the observed surround suppression in small-number tuning curves, tuning flank suppression is a known mechanism to contrast task-relevant and task-irrelevant stimulus features in attention- and working-memory-related operations[53–55]. Here as well, a suppressive zone below baseline is seen in the surround of the

https://doi.org/10.1038/s41562-023-01709-3

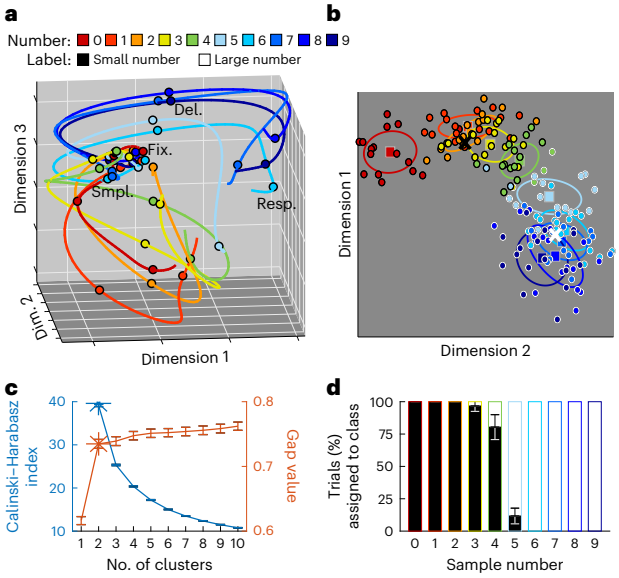


Fig. 5 | Population state-space analysis and *k*-means clustering. a, Averaged state-space trajectories of number-selective neurons for all number conditions, reduced to the three principal dimensions for visualization. Each trajectory depicts the temporal evolution in the time window −300 to 1,200 ms (stimulus onset to 100 ms after delay offset). The state space shows a gap between trajectories for numbers 0–4 versus 5–9. Circles indicate boundaries between task phases. Fix, Fixation; Smpl., Sample; Del., Delay; Resp., Response. **b**, Neural states, reduced to the two principal dimensions, after averaging firing rates per trial across the significant time window in the SVM classification analysis (60–1,200 ms). Different colours correspond to different number conditions. Each dot represents one trial; squares and ellipses indicate condition mean and covariance ellipse per condition. The colours of the dot outlines (black for 0–4 or white for 5–9) indicate the class label assigned by the *k*-means classifier. The black and grey crosses show the centroids of each class. **c**, Evaluation of different numbers of clusters using the Caliński–Harabasz criterion (blue) and the gap criterion (orange). Data are presented as mean values, error bars denote standard deviation (s.d.) of cross-validations ($n = 50$). Asterisks indicate the optimal number of clusters. Note that, unlike the Caliński–Harabasz criterion, the gap criterion would also be defined for clustering solutions containing only one cluster. **d**, Proportion of trials per number condition that were labelled as belonging to class 'small numbers' (black) or class 'large numbers' (white). Data are presented as mean values; error bars denote s.d. of cross-validations.

preferred stimulus parameter[56–59]. In intermediate-level visual area V4, surround suppression caused by spatial attention can be very quick and as early as 75 ms post stimulus onset[58]. The hypothesis is that with high attentional demand, the subitizing system overrides the estimation system. Thus, the subitizing system would enhance rather than replace estimation for small numbers[5]. This would also explain why estimation processes can in principle work also for small numbers, as seen many times in both human and animal brain studies[7,8,12,60–63]. This hypothesis is consistent with our current findings but needs to be tested empirically by contrasting the responses of neurons with and without attentional demands assigned to number representations. Complementing our parity judgement task with richer and more explicit number tasks could also help to support the generality of findings.

Methods

Experimental model and participant details

All studies conformed to the guidelines of the Medical Institutional Review Board at the University of Bonn, Germany, and were approved by this Board (licence no. 146/19). Seventeen human participants (five male, mean age 37.6 years) with medically refractory focal epilepsy undergoing invasive pre-surgical assessment participated in the study. Informed written consent was obtained from each patient; participants received no financial compensation for participating in the study.

Neurophysiological recording

Participants were implanted bilaterally with chronic intracerebral depth electrodes in the MTL to localize the seizure-onset zone for possible neurosurgical resection. The implantation site of the electrodes was determined exclusively by clinical criteria and varied across participants. To record neuronal signals, we used 9–10 clinical Behnke–Fried depth electrodes (AD-Tech Medical Instrument Corp.). Each depth electrode contained a bundle of nine platinum–iridium micro-wires protruding ~4 mm from the tip of each electrode: eight high-impedance active recording channels, and one low-impedance reference wire. Using a 256-channel ATLAS neurophysiology system (Neuralynx), differential neuronal signals (recording range ±3,200 µV) were filtered (bandwidth 0.1–9,000 Hz), amplified and digitized (sampling rate 32,768 Hz). Recorded spikes and behavioural data were synchronized via 64-bit timestamps using the Cheetah software (Neuralynx).

After bandpass-filtering (bandwidth 300–3,000 Hz) the local field potentials, spikes were automatically detected and pre-sorted using the Combinato package[64]. Classification as artefact, multi-unit or single unit was verified manually on the basis of spike shape and its variance, inter-spike-interval distribution per cluster, and the presence of a plausible refractory period. Only units that responded with an average firing rate of >1 Hz during stimulus presentation were included in the analyses. Across 28 recording sessions from all 17 participants, a total of 801 single units were identified in the PHC (109 units), EC (262 units), HIPP (275 units) and AMY (155 units).

Stimuli

All stimuli were presented within a filled grey circle (diameter approximately 6° of visual angle) on a black background. During fixation and delay phase, a white fixation spot was presented in the centre of the grey area. It was removed during stimulus presentation to avoid confusion with non-symbolic stimuli.

Numerical values of the stimuli ranged from 0 to 9 and consisted of black sets of dots with the number of dots corresponding to the respective numerical value ('numerosities'). Given that we needed zero for a balanced count of even and odd numbers, and acknowledging that zero is a set (even if empty) and a whole number like the natural numbers, we included zero in the stimulus presentation. We used different 'protocols' to control for low-level visual features. For the standard protocol, diameter and location of each dot varied randomly within a given range (diameters of 0.3° to 0.8° of visual angle). In the control displays, the total dot area and dot density (mean distances between centres of the dots) across numerosities was equated. Additionally, in half of the control trials, the dots were linearly arranged. Standard and control protocols for the non-symbolic stimuli were shown with equal probability of 50%.

Experimental task

Participants performed a parity judgement task sitting in bed and facing a laptop (display diagonal 11.7 inches, resolution 1,366 × 768 px) on which stimuli were presented at a distance of approximately 50 cm. Participants were not informed about hypotheses or purposes of the experiment to exclude any bias.

Before the experiment, the task instruction was displayed on the screen in addition to verbal explanation by the experimenter, specifying which numbers were 'even' and which ones 'odd'. Furthermore, to reduce confusion about the 'zero' stimulus, we added some familiarization trials preceding the recordings, during which the experimenter pointed out, once more, that an empty grey circle represented the 'even number zero'.

Each trial started with a fixation period of 300 ms. Afterwards, a number stimulus was presented for 500 ms, followed by a 600 ms delay display. After delay offset, participants had to decide whether the number had been even or odd by pressing the left or right arrow key on the keyboard, respectively, as indicated on the response screen ('gerade' (even) or 'ungerade' (odd)). To control for potential motor bias, the keys associated with the respective response were balanced and switched across blocks. Participants responded in a self-paced manner, but were asked to respond as fast and accurately as possible. After a 200 ms feedback display, the next trial started automatically. Each number stimulus was presented 16 times, resulting in 160 trials. A session was divided into four blocks, comprising all conditions in pseudo-random order. Stimuli and experimental task were programmed in MATLAB R2017a (The MathWorks), using the Psychtoolbox3 (refs. 65–67).

Behavioural analyses

First, we plotted the behavioural measures (error rates and RTs, averaged across participants, $n = 17$) as a function of numerical value of the stimulus. This function is characterized by a shallow, near-zero slope for small numbers, and a steeper slope for numbers beyond the subitizing range. The discontinuity point, in which the slope of this function changes, defines the upper boundary of the subitizing range. To quantify this boundary, we applied the algorithm for calculating individual subitizing ranges[18], that is, we first fitted a sigmoid (logistic) function to the behavioural data:

$$\mathrm{BM} = L + (U - L) \cdot \frac{1}{1 + \exp(-x - \mathrm{IP})}$$

where BM is the behaviour exhibited in response to the presentation of numerical value x. The model coefficients lower bound L, upper bound U and inflection point IP were estimated in the fitting process. We then applied the Levenberg–Marquardt algorithm to solve this non-linear least-squares curve-fitting problem. Next, we fitted two linear functions to the sigmoid curve. The subitizing line is equivalent to the lower bound L where the sigmoid curve crosses the y axis; the tangent line is fitted to the tangent at the inflection point IP. The intersection point of these two linear fits is then used as a proxy for the upper boundary of the subitizing range.

Neuronal analyses

Overall behavioural performance was high across all participants (mean ± standard deviation: 86.4 ± 3.1%). Errors occurred mainly for larger numbers. Because of the low error rate and the need to have balanced numbers of trials across numerosities, we included both correct and incorrect trials into the analyses.

Tuning characteristics

Spike trains were smoothed trial-wise (Gaussian kernel with $\sigma = 150$ ms) for each unit within the trial window −300 to 1,200 ms (fixation onset to 100 ms after delay offset). At every 20 ms step, instantaneous firing rates were subjected to a two-factor ANOVA with factors 'numerical value' (0–9) and 'protocol' (standard versus control) to detect tuning to numerical values, and a separate Mann–Whitney U test with factor 'parity' (even versus odd) to exclude neurons responsive to parity judgements (note that we could not apply a three-factor ANOVA as parity is not independent from the numerical value), resulting in a temporal sequence of P values for each of the three factors. A cluster permutation test[68] was then performed to identify time intervals of significant number encoding, thereby controlling for multiple comparisons across time ($\alpha_{clus} = 0.01$; $P_{rank} < 1\%$; $n_{perm} = 100$). A unit was termed 'exclusively number-selective' (NUM-ONLY) if a significant time interval for the factor 'numerical value' was observed between 0 ms and 1,000 ms (stimulus onset to 100 ms before delay offset), and there were no overlapping significant intervals for the factors 'protocol' or 'parity'. These units are henceforth referred to as 'number neurons'. Proportions of these number neurons were determined for each MTL region and subjected to a binomial test ($P_{chance} = 0.01$), Bonferroni-corrected for multiple comparisons across brain regions ($n = 4$), to evaluate whether the observed proportions were higher than expected by chance.

For each number neuron, we calculated tuning functions by averaging the firing rates during the significant time interval across trials for all numerical values. The numerical value eliciting the maximum response was defined as 'preferred numerosity'. To test for potential differences in the proportions of preferred numerosities, we applied the Mantel–Haenszel X^2 test[69,70], a generalized version of Pearson's X^2 test, for analysis of $2 \times 9 \times 17$ contingency tables, excluding zero as an outlier and stratified for different participants ($n = 17$).

Tuning functions were then standardized by z-scoring, that is, we subtracted the mean baseline activity elicited during the fixation periods (−300 to 0 ms) from all values and divided the difference by the standard deviation. In cases where multiple significant number intervals were identified within the same unit, we calculated separate tuning curves for each of these intervals (8/121 number neurons). Population tuning functions were then obtained by averaging across all units that preferred the same number.

To quantify surround suppression, we combined the firing rates to all non-preferred numbers for all units preferring the same numerical value and tested whether they were significantly smaller than spontaneous activity (that is, a z-score of 0) using a one-sided Wilcoxon signed rank test.

To estimate the tuning amplitude and width of all numerosity-selective neurons, we fitted a Gauss function, representing the standard symmetric distribution, to each individual tuning curve:

$$\mathrm{FR}(x) = a \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) + o$$

where FR is the z-scored firing rate elicited in response to the presentation of numerical value x. The mean μ was fixed to the preferred number; the model coefficients amplitude a, offset o, and standard deviation σ were estimated in the fitting process, thereby using the following bound constraints: $a = [0;\max(\mathrm{FR})]$, $o = [\min(\mathrm{FR});\max(\mathrm{FR})]$ and $\sigma = [0;\mathrm{Inf}]$, to avoid implausible fitting results. We then applied the Levenberg–Marquardt algorithm to solve this non-linear least-squares curve-fitting problem.

RSA

Pearson's correlation coefficient quantifies the strength of the linear relationship between two variables. To evaluate firing rate differences between different number conditions, we performed an RSA. We calculated a correlation matrix, showing the correlations between firing rates in response to number i and to number j, respectively, for all condition pairs, based on the z-scored tuning curves of all number neurons. We then determined the boundary that divided the data best into the categories of 'small' and 'large' numbers. For this, we defined eight potential boundaries (1|2, 2|3, …, 8|9). Due to obvious dissimilarity, zero was excluded from this categorization analysis. A boundary divided the correlation matrix into four squares. Correlation coefficients in the upper-left (within small-number category) and lower-right squares (within large-number category) of the matrix were then iteratively compared with the coefficients in the remaining upper-right and lower-left (across-category) matrix squares for different number boundaries. For each boundary, the difference between within-category and across-category elements was then quantified using a two-sided Mann–Whitney U test ($\alpha = 0.01$, Bonferroni-corrected for multiple comparisons of boundaries, $n = 8$). Note that the main diagonal was excluded as it reflects the correlation of each stimulus with itself, and that the correlation matrix is symmetric. Thus, only

values from the upper triangular portion of the correlation matrix were considered. For each boundary, we assessed the *P* value of the statistical test, and the average difference between within-category and across-category elements to evaluate how well the boundary divided the data into two categories. Pairs of correlation coefficients are not statistically independent due to their bivariate nature, which might bias the results of the Mann–Whitney *U* test. To account for this, we performed an additional permutation test by randomly shuffling the within-category and across-category labels 10,000 times and comparing the test statistics of the random data to the true ones, using again a two-sided Mann–Whitney *U* test[71].

SVM classification

For each unit, data were divided into ten classes according to the numerical value of the sample stimulus (16 trials per class) and spike trains per trial were smoothed (Gaussian kernel with $\sigma = 150$ ms) within the analysis window −300 to 1,200 ms (fixation onset to 100 ms after delay offset). A default multi-class SVM classifier[53] was then trained and tested on the instantaneous firing rates at every 20 ms step[72]. We applied Monte-Carlo cross-validation, that is, we created multiple splits of our dataset ($n_{\text{repetitions}} = 100$) by randomly sampling 50% of the trials as training set, balancing conditions within each split. The remaining 50% of the trials were used as test set. Thus, each training and test set comprised 80 trials. For each split, we standardized all firing rates by *z*-scoring (mean and standard deviation obtained from training data only), fitted the classifier to the training data, and assessed predictive accuracy by counting the instances for which a certain activity pattern of the test data was labelled correctly. The results were then averaged over all splits. To identify temporal clusters during which accuracy differed significantly from chance level (10% for ten classes), the analysis was repeated with randomly shuffled trial labels ($n_{\text{perm}} = 100$), and a cluster permutation test[63] was performed. In short, we identified temporal clusters of significant values by comparing the true accuracy values against the distribution of random ones ($\alpha_{\text{clus}} = 0.01$). The significance of these 'candidate clusters' was then evaluated by comparison with the clusters of the random data ($P_{\text{rank}} < 1\,\%$). Next, an SVM (with the same settings as above) was trained and tested on the firing rates obtained by averaging across the significant time window (60–1,200 ms). We assembled a confusion matrix, which counted the frequency at which a trial of a certain class was assigned different labels by the classifier.

Again, we analysed which boundary divided the data best into the categories of 'small' versus 'large' numbers. As before, we defined eight potential boundaries (1|2 to 8|9; excluding zero). Classification probabilities in the upper-left (within small-number category) and lower-right squares of the matrix (within large-number category) were iteratively compared with the classification probabilities in the remaining (across-category) matrix squares (lower left and upper right) for all number boundaries. The difference between both groups was then quantified using a non-parametric Mann–Whitney *U* test ($\alpha = 0.01$; Bonferroni-corrected for multiple comparisons of boundaries, $n = 8$). Note that the main diagonal was excluded as it reflects correct classifications, and that the confusion matrix is not symmetric (unlike the correlation matrix). For each boundary, we assessed the *P* value of the statistical test and the average difference between within-category and across-category elements to evaluate how well the boundary divided the data into two categories. The results of the Mann–Whitney *U* test were again verified using a permutation test.

Multi-dimensional state-space analysis

To analyse neural activity of a neuronal population, we defined an *n*-dimensional space, where each axis represents the instantaneous firing rate of a number-selective neuron. At any given time, the population activity is then characterized by a single point in this space, resulting in a neural trajectory as the activity evolves over time. In other words, we calculated the trajectories for the ten different numerosities in a 121-dimensional space after averaging across conditions and smoothing (Gaussian kernel with $\sigma = 150$ ms) spike trains for each number neuron. A Gaussian-process factor analysis model was then applied[73], and the resulting neural trajectories were orthonormalized to order the dimensions by the amount of data covariance they explain. For visualization, only the top three dimensions (in terms of covariance explained) were considered.

Next, the neural population state was calculated by averaging firing rates across the significant time window (60–1,200 ms) for all trials. The neural state space was then orthonormalized using principal component analysis. For visualization, only the top two dimensions were displayed. We then applied unsupervised *k*-means clustering to partition all trials ($n = 160$) into two clusters[33]. In short, *k*-means clustering iteratively partitions the data into *k* distinct non-overlapping clusters such that the distance between all elements of the cluster and every cluster's centroid is minimized. We used the squared Euclidean distance as a distance metric, that is, centroids are the arithmetic mean of the elements in that cluster and repeated the algorithm 50 times with new randomly chosen initial cluster centroid positions.

We applied two criteria to evaluate the optimal number of clusters for our data. First, we calculated the Caliński–Harabasz index[31], also called VRC, which is defined as the ratio between overall between-cluster variance and overall within-cluster variance. Maximizing the VRC value with respect to *k* classes yields the optimal number of classes. As a second criterion, we calculated the gap value[32]. It formalizes the heuristic 'elbow method', according to which the optimal number of clusters can be found by locating the most dramatic decrease in error measurement (the 'elbow' or 'gap'). Note that, unlike the Caliński–Harabasz criterion, the gap criterion is also defined for clustering solutions containing only one cluster. For cross-validation, the *k*-means clustering analysis was repeated 50 times, using only 75% randomly selected trials per condition for each cross-validation run.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data associated with this study are publicly available at https://github.com/EstherKutter/Distinct-Neuronal-Representation-Of-Small-And-Large-Numbers-In-The-Human-MTL.

Code availability

The custom code associated with this study is publicly available at https://github.com/EstherKutter/Distinct-Neuronal-Representation-Of-Small-And-Large-Numbers-In-The-Human-MTL.

References

1. Jevons, W. S. The power of numerical discrimination. *Nature* **3**, 281–282 (1871).
2. Kaufman, E. L., Lord, M. W., Reese, T. W. & Volkmann, J. The discrimination of visual number. *Am. J. Psychol.* **62**, 498–525 (1949).
3. Mandler, G. & Shebo, B. J. Subitizing: an analysis of its component processes. *J. Exp. Psychol. Gen.* **111**, 1–22 (1982).
4. Feigenson, L., Dehaene, S. & Spelke, E. Core systems of number. *Trends Cogn. Sci.* **8**, 307–314 (2004).
5. Anobile, G., Cicchini, G. M. & Burr, D. C. Number as a primary perceptual attribute: a review. *Perception* **45**, 5–31 (2016).
6. Cheyette, S. J. & Piantadosi, S. T. A unified account of numerosity perception. *Nat. Hum. Behav.* **4**, 1265–1272 (2020).
7. Tsouli, A. et al. The role of neural tuning in quantity perception. *Trends Cogn. Sci.* **26**, 11–24 (2022).
8. Piazza, M., Mechelli, A., Butterworth, B. & Price, C. J. Are subitizing and counting implemented as separate or functionally overlapping processes? *NeuroImage* **15**, 435–446 (2002).

9. Libertus, M. E., Woldorff, M. G. & Brannon, E. M. Electrophysiological evidence for notation independence in numerical processing. *Behav. Brain Funct.* **3**, 1 (2007).
10. Harvey, B. M., Klein, B. P., Petridou, N. & Dumoulin, S. O. Topographic representation of numerosity in the human parietal cortex. *Science* **341**, 1123–1126 (2013).
11. Fornaciai, M. & Park, J. Decoding of electroencephalogram signals shows no evidence of a neural signature for subitizing in sequential numerosity. *J. Cogn. Neurosci.* **33**, 1535–1548 (2021).
12. Cai, Y. et al. Topographic numerosity maps cover subitizing and estimation ranges. *Nat. Commun.* **12**, 3374 (2021).
13. Sathian, K. et al. Neural evidence linking visual object enumeration and attention. *J. Cogn. Neurosci.* **11**, 36–51 (1999).
14. Fink, G. R. et al. Deriving numerosity and shape from identical visual displays. *NeuroImage* **13**, 46–55 (2001).
15. Hyde, D. C. & Spelke, E. S. All numbers are not equal: an electrophysiological investigation of small and large number representations. *J. Cogn. Neurosci.* **21**, 1039–1053 (2009).
16. Kutter, E. F., Bostroem, J., Elger, C. E., Mormann, F. & Nieder, A. Single neurons in the human brain encode numbers. *Neuron* **100**, 753–761.e4 (2018).
17. Kutter, E. F., Boström, J., Elger, C. E., Nieder, A. & Mormann, F. Neuronal codes for arithmetic rule processing in the human brain. *Curr. Biol.* **32**, 1275–1284.e4 (2022).
18. Leibovich-Raveh, T., Lewis, D. J., Kadhim, S. A. R. & Ansari, D. A new method for calculating individual subitizing ranges. *J. Numer. Cogn.* **4**, 429–447 (2018).
19. Atkinson, J., Campbell, F. W. & Francis, M. R. The magic number 4±0: a new look at visual numerosity judgements. *Perception* **5**, 327–334 (1976).
20. Simon, T. J. & Vaishnavi, S. Subitizing and counting depend on different attentional mechanisms: evidence from visual enumeration in afterimages. *Percept. Psychophys.* **58**, 915–926 (1996).
21. Sengupta, R., Surampudi, B. R. & Melcher, D. A visual sense of number emerges from the dynamics of a recurrent on-center off-surround neural network. *Brain Res.* **1582**, 114–124 (2014).
22. Sengupta, R., Bapiraju, S. & Melcher, D. Big and small numbers: empirical support for a single, flexible mechanism for numerosity perception. *Atten. Percept. Psychophys.* **79**, 253–266 (2017).
23. Fias, W. Two routes for the processing of verbal numbers: evidence from the SNARC effect. *Psychol. Res.* **65**, 250–259 (2001).
24. Nuerk, H.-C., Iversen, W. & Willmes, K. Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Q. J. Exp. Psychol. A* **57**, 835–863 (2004).
25. Nieder, A. Representing something out of nothing: the dawning of zero. *Trends Cogn. Sci.* **20**, 830–842 (2016).
26. Schoups, A., Vogels, R., Qian, N. & Orban, G. Practising orientation identification improves orientation coding in V1 neurons. *Nature* **412**, 549–553 (2001).
27. Yang, T. & Maunsell, J. H. R. The effect of perceptual learning on neuronal responses in monkey visual area V4. *J. Neurosci.* **24**, 1617–1626 (2004).
28. Lee, S. H. et al. Activation of specific interneurons improves V1 feature selectivity and visual perception. *Nature* **488**, 379–383 (2012).
29. Nieder, A. & Miller, E. K. Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron* **37**, 149–157 (2003).
30. Saxena, S. & Cunningham, J. P. Towards the neural population doctrine. *Curr. Opin. Neurobiol.* **55**, 103–111 (2019).
31. Caliński, T. & Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.* **3**, 1–27 (1974).
32. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. Ser. B* **63**, 411–423 (2001).
33. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **28**, 129–137 (1982).
34. Viswanathan, P. & Nieder, A. Differential impact of behavioral relevance on quantity coding in primate frontal and parietal neurons. *Curr. Biol.* **25**, 1259–1269 (2015).
35. Ramirez-Cardenas, A., Moskaleva, M. & Nieder, A. Neuronal representation of numerosity zero in the primate parieto-frontal number network. *Curr. Biol.* **26**, 1285–1294 (2016).
36. Viswanathan, P. & Nieder, A. Spatial neuronal integration supports a global representation of visual numerosity in primate association cortices. *J. Cogn. Neurosci.* **32**, 1184–1197 (2020).
37. Hartline, H. K., Wagner, H. G. & Ratliff, F. Inhibition in the eye of Limulus. *J. Gen. Physiol.* **39**, 651–673 (1956).
38. Isaacson, J. S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).
39. Diester, I. & Nieder, A. Complementary contributions of prefrontal neuron classes in abstract numerical categorization. *J. Neurosci.* **28**, 7737–7747 (2008).
40. Ditz, H. M., Fechner, J. & Nieder, A. Cell-type specific pallial circuits shape categorical tuning responses in the crow telencephalon. *Commun. Biol.* **5**, 269 (2022).
41. Angelucci, A. & Bressloff, P. C. Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Prog. Brain Res.* **154**, 93–120 (2006).
42. Bair, W., Cavanaugh, J. R. & Movshon, J. A. Time course and time–distance relationships for surround suppression in macaque V1 neurons. *J. Neurosci.* **23**, 7690–7701 (2003).
43. Nassi, J. J., Lomber, S. G. & Born, R. T. Corticocortical feedback contributes to surround suppression in V1 of the alert primate. *J. Neurosci.* **33**, 8504–8517 (2013).
44. Ison, M. J. et al. Selectivity of pyramidal cells and interneurons in the human medial temporal lobe. *J. Neurophysiol.* **106**, 1713–1721 (2011).
45. Gast, H. et al. Burst firing of single neurons in the human medial temporal lobe changes before epileptic seizures. *Clin. Neurophysiol.* **127**, 3329–3334 (2016).
46. Mosher, C. P. et al. Cellular classes in the human brain revealed in vivo by heartbeat-related modulation of the extracellular action potential waveform. *Cell Rep.* **30**, 3536–3551.e6 (2020).
47. Railo, H., Koivisto, M., Revonsuo, A. & Hannula, M. M. The role of attention in subitizing. *Cognition* **107**, 82–104 (2008).
48. Vetter, P., Butterworth, B. & Bahrami, B. Modulating attentional load affects numerosity estimation: evidence against a pre-attentive subitizing mechanism. *PLoS ONE* **3**, e3269 (2008).
49. Burr, D. C., Turi, M. & Anobile, G. Subitizing but not estimation of numerosity requires attentional resources. *J. Vis.* **10**, 20 (2010).
50. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
51. Cowan, N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* **24**, 87–114 (2001).
52. Piazza, M., Fumarola, A., Chinello, A. & Melcher, D. Subitizing reflects visuo-spatial object individuation capacity. *Cognition* **121**, 147–153 (2011).
53. Wang, X.-J., Tegnér, J., Constantinidis, C. & Goldman-Rakic, P. S. Division of labor among distinct subtypes of inhibitory neurons in a cortical microcircuit of working memory. *Proc. Natl Acad. Sci. USA* **101**, 1368–1373 (2004).
54. Störmer, V. S. & Alvarez, G. A. Feature-based attention elicits surround suppression in feature space. *Curr. Biol.* **24**, 1985–1988 (2014).
55. Kiyonaga, A. & Egner, T. Center-surround inhibition in working memory. *Curr. Biol.* **26**, 64–68 (2016).

56. Müller, N. G. & Kleinschmidt, A. The attentional 'spotlight's' penumbra: center-surround modulation in striate cortex. *NeuroReport* **15**, 977–980 (2004).
57. Hopf, J.-M. et al. Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proc. Natl Acad. Sci. USA* **103**, 1053–1058 (2006).
58. Sundberg, K. A., Mitchell, J. F. & Reynolds, J. H. Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron* **61**, 952–963 (2009).
59. Anton-Erxleben, K., Stephan, V. M. & Treue, S. Attention reshapes center-surround receptive field structure in macaque cortical area MT. *Cereb. Cortex* **19**, 2466–2478 (2009).
60. Nieder, A., Freedman, D. J. & Miller, E. K. Representation of the quantity of visual items in the primate prefrontal cortex. *Science* **297**, 1708–1711 (2002).
61. Nieder, A. & Merten, K. A labeled-line code for small and large numerosities in the monkey prefrontal cortex. *J. Neurosci.* **27**, 5986–5993 (2007).
62. Ditz, H. M. & Nieder, A. Neurons selective to the number of visual items in the corvid songbird endbrain. *Proc. Natl Acad. Sci. USA* **112**, 7827–7832 (2015).
63. Ditz, H. M. & Nieder, A. Format-dependent and format-independent representation of sequential and simultaneous numerosity in the crow endbrain. *Nat. Commun.* **11**, 686 (2020).
64. Niediek, J., Boström, J., Elger, C. E. & Mormann, F. Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours. *PLoS ONE* **11**, e0166598 (2016).
65. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
66. Pelli, D. G. The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat. Vis.* **10**, 437–442 (1997).
67. Kleiner, M., Brainard, D. & Pelli, D. What's new in Psychtoolbox-3? *Perception* **36**, 14 (2007).
68. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
69. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl Cancer Inst.* **22**, 719–748 (1959).
70. Somes, G. W. The generalized Mantel–Haenszel statistic. *Am. Stat.* **40**, 106–108 (1986).
71. Mormann, F. et al. Neurons in the human amygdala encode face identity, but not gaze direction. *Nat. Neurosci.* **18**, 1568–1570 (2015).
72. Chang, C.-C. & Lin, C.-J. LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **102**, 1–27 (2011).
73. Yu, B. M. et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–635 (2009).

Acknowledgements

We thank all patients for their participation. This research was supported by the German Research Council (Mo 930/4-2, SPP 2205, SPP 2411, SFB 1089; Ni 618/11-1, SPP 2205), the BMBF (031L0197B) and a NRW Network Grant (iBehave). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

A.N. and F.M. designed the study; R.S. and F.M. recruited patients; V.B. and F.M. implanted the electrodes; E.F.K. and G.D. collected the data; E.F.K. and A.N. analysed the data with contributions from F.M.; A.N., E.F.K. and F.M. wrote the paper. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-023-01709-3.

Correspondence and requests for materials should be addressed to Florian Mormann or Andreas Nieder.

Peer review information *Nature Human Behaviour* thanks Rakesh Sengupta and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): Andreas Nieder

Last updated by author(s): Aug 18, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Depth electrodes were furnished with bundles of nine microwires each (eight high-impedance recording electrodes, one low-impedance reference, AdTech, Racine, WI) protruding ~4mm from the electrode tips. Recording system: Neuralynx ATLAS system (Bozeman, MT), Neuralynx Cheetah software (Bozeman, MT) Spike-sorting: Combinato package, Version 1 (Niediek et al., PLOS ONE 11 (12): e0166598. 2016) (https://github.com/jniediek/combinato) Stimulus presentation: Psychtoolbox3 (www.psychtoolbox.org)
Data analysis	Matlab R2017a (The MathWorks Inc., Natick, MA)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Data and analysis scripts to reproduce the main results can be found in a GitHub repository (https://github.com/EstherKutter/Distinct-Neuronal-Representation-Of-Small-And-Large-Numbers-In-The-Human-MTL)

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Both sexes participated randomly in the study (five males, twelve females).
Population characteristics	All participants were right-handed. No population characteristics were included as covariates in the current study.
Recruitment	The current sample is an opportunity sample, based on patient influx for epilepsy treatment. Patients were included in the study if they consented to microwire implantation.
Ethics oversight	Informed written consent was obtained from each patient. All studies conformed to the guidelines of the Ethik-Kommission, Medizinische Fakultät, Rheinische Friedrich-Wilhelms-Universität Bonn, i.e. in English: the Medical Institutional Review Board at the University of Bonn, Germany (license Nr. 146/19).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Seventeen human subjects (five males, twelve females, mean age 37.6 years) with medically refractory focal epilepsy undergoing invasive presurgical assessment participated in the study. The sample is representative of patients with intractable epilepsy. No a-priori sample-size calculation was performed. The number of recording sessions comfortably complies with or exceeds current standards in human microwire recordings.
Data exclusions	no data were excluded
Replication	not applicable due to the constraints of the intervention
Randomization	there was no allocation to experimental groups
Blinding	blinding was not possible due to the constraints of the intervention

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

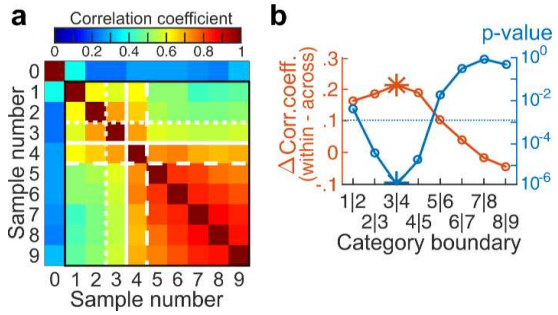
Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

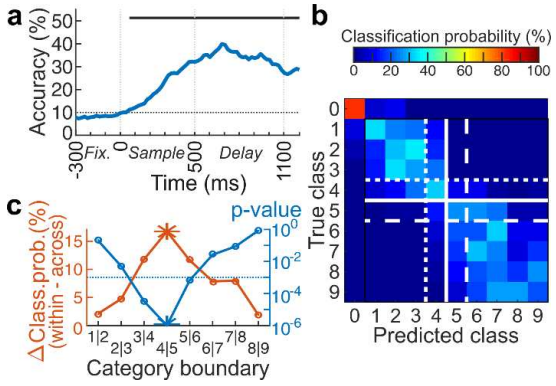
Supplementary Information



Supplementary Figure 1: Related to Figure 3. ***Representational similarity analysis for the entire population of MTL single units.***

a) Correlation coefficients of the z-scored firing rates across the entire population of units (n = 801) between pairs of numbers. Firing rates were more similar (higher correlations, corresponding to reddish colors) for numbers from the same number category (small or large, upper-left and lower-right square, respectively), compared to responses for numbers from different categories (lower-left and upper-right square). White lines depict significant number category boundaries (solid line is most significant), dividing correlations into small versus large number categories.

b) Evaluation of the goodness-of-fit of different number boundaries. Orange values depict the differences of correlation coefficients when segregating small versus large number categories (excluding zero) at different boundaries. The corresponding p-values (two-sided Mann-Whitney U-test) for these coefficient differences are shown in blue. Boundary 3 versus 4 (asterisks) divides the data most significantly into two number categories. The blue dotted line indicates α = 0.01, Bonferroni-corrected for multiple comparisons (n = 8).

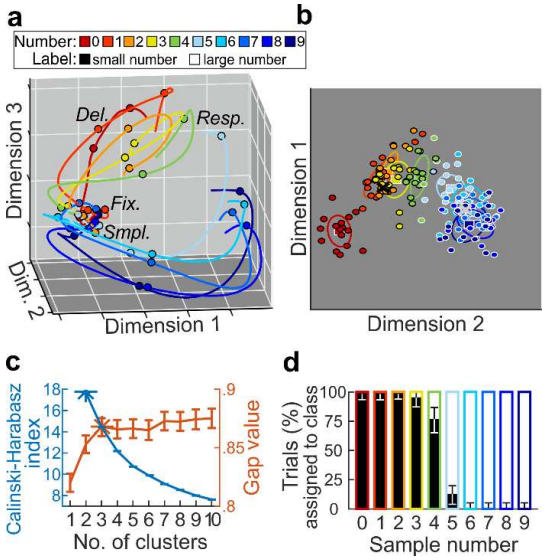


Supplementary Figure 2: Related to Figure 4: *Support vector machine (SVM) classification analysis for the entire population of MTL single units.*

a) Classification accuracy for decoding number information after training an SVM classifier on the instantaneous firing rates across the trial period for all neurons *(n* = 801). The dashed line represents chance level (10 % for ten classes). The black bar above the data indicates significance ($p < 0.01$, one-sided permutation test compared to SVM trained on shuffled data).

b) Confusion matrix derived from training an SVM classifier on firing rates averaged across the significant time window in the sliding-window analysis in **a** (60–1200 ms). White lines depict the significant boundaries (highest significance for the solid, thick line) that divide the number range into small and large number categories.

c) Evaluation of the goodness of different number boundaries. Orange values depict the difference in classification probabilities when segregating small versus large number categories at different boundaries (excluding zero). The corresponding *p*-values (two-sided Mann-Whitney U-test) for these probability differences are shown in blue. Boundary 4 versus 5 (asterisks) divides the data most significantly into two number categories. The blue dotted line indicates $\alpha = 0.01$, Bonferroni-corrected for multiple comparisons ($n = 8$).



Supplementary Figure 3: Related to Figure 5: *Population state space analysis and k-means clustering for the entire population of MTL single units.*

a) Averaged state-space trajectories of all neurons for all number conditions, reduced to the three principal dimensions for visualization. Each trajectory depicts the temporal evolution in the time window -300–1200 ms (stimulus onset to 100 ms after delay offset). The state-space shows a gap between trajectories for numbers 0–4 versus 5–9. Circles indicate boundaries between task phases.

b) Neural states, reduced to the two principal dimensions, after averaging firing rates per trial across the significant time window in the SVM classification analysis (60–1200 ms). Different colors correspond to different number conditions. Each dot represents one trial; squares and ellipses indicate condition mean and covariance ellipse per condition. The colors of the dot outlines (black for 0–4 or white for 5–9) indicate the class label assigned by the *k*-means classifier. The black and gray crosses show the centroids of each class.

c) Evaluation of different numbers of clusters using the Caliński-Harabasz criterion (blue) and the gap criterion (orange). Data are presented as mean values, error bars denote STD of cross-validations ($n = 50$). Asterisks indicate the optimal number of clusters. Note that, unlike the Caliński-Harabasz criterion, the gap criterion would also be defined for clustering solutions containing only one cluster.

d) Proportion of trials per number condition that were labelled as belonging to class ‘small numbers’ (black) or class ‘large numbers’ (white). Data are presented as mean values, error bars denote STD of cross-validations.